非数学类专业
研究生教学用书

非线性最优化计算方法

Computational Methods for Nonlinear Optimization

主　编　张光澄
副主编　王文娟　韩会磊　张　雷

高等教育出版社

ISBN 7-04-016699-2
9 787040 166996 >
定价 36.20 元

非数学类专业
研究生教学用书

非线性最优化计算方法

Computational Methods for Nonlinear Optimization

主　编　张光澄
副主编　王文娟　韩会磊　张　雷

高等教育出版社

内容简介

本书是作者多年来为四川某些高校理工科相关专业的硕士研究生、工程硕士生、本科生开设“最优化方法”课程，在教学实践及所编教材的基础上，对搜集整理的大量材料经过充分酝酿，反复修订而成的。

本书在课程内容处理上遵循如下原则：突出方法，注重概念，适当介绍算法的基本理论；强调应用，加强算法实现的基本训练；引导学生主动思考，激发学生的学习兴趣；通过算法到程序设计有序而系统的训练，提高学生程序设计的能力。

全书分为上、下两篇，上篇共9章，介绍无约束最优化方法。第1章为基础知识，介绍凸集的基本性质，函数及凸函数的最优性条件；第2章综述最优化问题及无约束最优化算法；第3章至第9章分别介绍求解无约束最优化问题的各种算法。下篇共9章，介绍约束最优化方法。第10、11两章介绍线性规划问题及其解法；第12章至16章介绍非线性规划的最优性条件及常用的算法；第17章讨论离散系统的动态规划方法；第18章简单介绍了包括禁忌搜索、模拟退火、遗传算法和神经网络算法等现代优化算法。

本书可作为高等院校理工科相关专业的硕士生、工程硕士生以及高年级本科生的教材，也可供从事相关工作的科研人员和工程技术人员参考。

图书在版编目(CIP)数据

非线性最优化计算方法/张光澄主编. —北京：高等教育出版社，2005.7
ISBN 7-04-016699-2

Ⅰ.非... Ⅱ.张... Ⅲ.非线性-最优化算法-高等学校-教材 Ⅳ.O224

中国版本图书馆CIP数据核字（2005）第043761号

策划编辑 李艳馥 **责任编辑** 舒放江 **封面设计** 李卫青
责任绘图 郝 林 **版式设计** 史新薇 **责任校对** 杨雪莲
责任印制 杨 明

出版发行	高等教育出版社	购书热线	010-58581118
社 址	北京市西城区德外大街4号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经 销	北京蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印 刷	国防工业出版社印刷厂		
开 本	787×960 1/16	版 次	2005年7月第1版
印 张	26.5	印 次	2005年7月第1次印刷
字 数	450 000	定 价	36.20元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

物料号 16699-00

前　言

最优化方法和理论来源于军事、管理、经济和工程技术领域的各个方面，其内容的深度和广度也随着各个不同阶段的科学技术水平而发展。两次世界大战，尤其是二战中提出了很多军事最优化问题。这些问题及其解决方法具有鲜明的特色：数据是实践中的真实数据，解决问题的人员组成是多学科的，处理问题的方法渗透着物理学的思想。

第二次世界大战以后最优化方法的应用由军事问题转入民用问题，提出了工程技术和现代管理的最优化理论和方法。特别是近三十多年来，由于科学技术的需要以及计算机和计算技术的发展，为求解各种复杂问题的最优化方法和理论提供了雄厚的基础和手段。最优化的应用范围也愈来愈广，涉及设计、操作、工业过程，生产装置的分析以及生产计划制定，经济运作的决策等等问题。由于最优化方法是寻找最好效果的条件的方法，所以具有十分重要的现实意义。

作者多年来为四川大学等高校理工科相关专业的硕士研究生、高年级本科生开设"最优化计算方法"课程。本书系统介绍非线性最优化计算方法及理论，是在教学实践及所编教材的基础上反复修订而成的。在内容的处理上，遵循如下原则：注重概念、突出方法，适当介绍方法的基本理论；加强算法实现的基本训练，通过算法到程序设计有序而系统的训练，提高学生数值计算和程序设计的能力；突出应用，结合实际问题介绍求解数学模型的常用算法，并配有数值计算实例，吸引学生思考，激发其学习兴趣。

全书分上、下两篇，共计 18 章。上篇第 1 章至第 9 章介绍无约束最优化方法。第 1 章介绍最优化方法的基础知识，包括凸函数的概念及基本性质，函数及凸函数的最优性条件；第 2 章概述最优化问题及无约束最优化算法；第 3 章至第 9 章分别介绍常用的无约束最优化方法，如一维寻查方法、Newton 方法及拟 Newton 方法、共轭方向法、直接方法、线性最小二乘法及非线性最小二乘法等等。

下篇第 10 章至第 18 章介绍约束最优化方法。第 10 章至第 11 章介绍线性规划及整数规划；第 12 章至第 13 章介绍非线性规划的最优性条件和对偶理论；第 14 章至第 16 章介绍常用的约束最优化算法，包括可行方向法、投影梯度法、罚函数及精确罚函数法、二次规划方法等等；第 17

章介绍离散系统的动态规划方法；第18章介绍处理组合优化问题的现代优化方法。

本书的附录Ⅱ介绍了最优化方法产生的历史背景和发展进程，附录Ⅲ介绍当今世界上最流行的数学软件Matlab及其在最优化方法、数据处理方面的应用。

本书编写工作具体分工如下：

上卷第1章至第9章及全书的三个附录由张光澄、张雷联合执笔；下卷第10章、第11章、第18章由王文娟执笔；第12章、第14章、第15章、第16章由韩会磊执笔；第13章由范安东执笔；第17章由张光澄执笔。全书由张光澄、王文娟统稿完成编写工作。

作者十分感谢高等教育出版社对本书编辑与出版所给予的关心和支持。感谢四川大学王荫清教授、王宝富副院长和成都理工大学施泽进教授、魏贵民教授、郭科院长对本书编写工作给予的关心和帮助。四川大学侯泽华副教授和成都理工大学的张章华、田宝单、叶翠莲、冉伟等同学帮助整理本书初稿做了大量细致的工作，在此一并致以最诚挚的谢意。

作者力图使本书反映最优化方法的多样性、基本理论的严密性、应用的广泛性，但由于作者水平有限，在思想的表达和对内容的取舍、安排、衔接等方面难免存在缺点甚至错误，恳请读者批评指正。

张光澄

2005年元月

本书主要符号索引

$\mathbf{R}^n$——n 维实向量空间

T——向量或矩阵的转置

$\boldsymbol{x},\boldsymbol{x}^*,\boldsymbol{x}_k(k=1,2,\cdots)$——$n$ 维实列向量，记为

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$$

$$\boldsymbol{x}^*=(x_1^*,x_2^*,\cdots,x_n^*)^{\mathrm{T}}$$

$$\boldsymbol{x}_k=(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)})^{\mathrm{T}}$$

$\boldsymbol{e}_i$——第 i 个分量为 1 的单位向量

$\forall \boldsymbol{x}\in\mathbf{R}^n$——属于 $\mathbf{R}^n$ 的任意 n 维实向量

$f:S\subset\mathbf{R}^n\to\mathbf{R}^1$——$f$ 是定义在 $\mathbf{R}^n$ 中区域 S 上的实值函数

$\boldsymbol{A},\boldsymbol{B},\boldsymbol{H},\boldsymbol{U},\boldsymbol{I},\boldsymbol{G}$ 等——矩阵

$\|\cdot\|$——向量或矩阵的范数（$\|\cdot\|_2$ 为欧氏范数，$\|\cdot\|_\infty$ 为无穷范数）

$\det(\boldsymbol{A})$——矩阵 $\boldsymbol{A}$ 的行列式

$\operatorname{rank}(\boldsymbol{A})$——矩阵 $\boldsymbol{A}$ 的秩

$[\boldsymbol{A}|\boldsymbol{b}]$——线性方程组的系数矩阵 $\boldsymbol{A}$ 与右端向量 $\boldsymbol{b}$ 形成的增广矩阵

$\approx$——近似于

$\in$——属于

$\notin$——不属于

$\cup$——集合的并运算

$\cap$——集合的交运算

$\supset$——集合之间的包含

$\subset$——集合之间的被包含

$\exists$——存在

s.t.——“满足于”，“Subject to”的缩写

$N_\varepsilon(\bar{\boldsymbol{x}})$——以 $\bar{\boldsymbol{x}}$ 为中心，ε 为半径的邻域

$N(\boldsymbol{x}_0,\delta)$——以 $\boldsymbol{x}_0$ 为中心，δ 为半径的邻域

$\nabla f(\boldsymbol{x})$——$f(\boldsymbol{x})$ 的梯度，或记为 $\boldsymbol{g}(\boldsymbol{x})=\nabla f(\boldsymbol{x})$

$\nabla^2 f(\boldsymbol{x})$——$f(\boldsymbol{x})$ 的 Hessan 梯度，或记为 $\boldsymbol{H}(\boldsymbol{x})=\nabla^2 f(\boldsymbol{x})$

epi $f(\boldsymbol{x})$——$f(\boldsymbol{x})$ 的上图像，简称 $f(\boldsymbol{x})$ 的上图

hyp $f(\boldsymbol{x})$——$f(\boldsymbol{x})$ 的下图像，简称 $f(\boldsymbol{x})$ 的下图

int S——集合 S 的内部

H-终止条件——D. M. Himmeiblan 提出的判别极小点的收敛准则

$\boldsymbol{A}$-内积——两向量关于矩阵 $\boldsymbol{A}$ 的内积，即 $(\boldsymbol{x},\boldsymbol{y})_A=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$

目　　录

上篇　无约束最优化方法

第 1 章　基础知识 …… 1

1.1　凸集及其基本性质 …… 1

1.2　极值(一般函数)的最优性条件 …… 6

1.2.1　多元函数极值概念 …… 6

1.2.2　梯度与 Hesse 矩阵 …… 7

1.2.3　局部极值的最优性条件 …… 11

1.3　凸函数及凸函数极值的最优性条件 …… 14

1.3.1　凸函数的定义及判定 …… 14

1.3.2　凸函数的次梯度 …… 17

1.3.3　凸函数的最优性条件 …… 19

1.4　拟凸函数与全局最优 …… 21

第 2 章　最优化方法概述 …… 25

2.1　最优化问题的提法及分类 …… 25

2.2　最优化问题举例 …… 26

2.3　无约束极值问题算法综述 …… 30

2.3.1　下降算法 …… 31

2.3.2　算法收敛速度及终止法则 …… 35

2.3.3　收敛性条件 …… 39

第 3 章　一维搜索(寻查) …… 42

3.1　搜索(寻查)区间的确定 …… 43

3.2　二分法 …… 45

3.3　直接方法 …… 46

3.3.1　0.618 法(黄金分割法) …… 47

3.3.2　分数法(Fibonacci 法) …… 49

第 4 章　Newton 方法及其改进 …… 54

4.1　Newton 方法及其局限性 …… 54

4.2　Newton 算法的改进 …… 56

4.3　特征值法(Greenstadt 方法) …… 58

4.4　Newton 算法的 Gill 和 Murray 修正方案 …… 60

第 5 章　共轭方向法 …… 64

5.1 共轭方向 …… 64
5.2 共轭方向法 …… 66
5.3 共轭梯度法 …… 71
5.3.1 正定二次函数的基本算法 …… 71
5.3.2 基本性质 …… 74
5.3.3 一般函数的共轭梯度法 …… 75
第6章 拟 Newton 法 …… 79
6.1 尺度矩阵意义下的最速下降方法 …… 79
6.2 DFP 公式及 DFP 算法 …… 81
6.2.1 DFP 公式及其基本性质 …… 81
6.2.2 DFP 算法 …… 83
6.2.3 DFP 算法的二次收敛性质 …… 85
6.3 DFP 对偶公式及其等价形式 …… 88
6.3.1 DFP 对偶公式及其基本性质 …… 88
6.3.2 DFP 对偶公式的几种等价形式 …… 89
6.4 DFP 对偶算法 …… 91
6.4.1 修正矩阵 B_{k+1} 的 LDL^T 分解 …… 92
6.4.2 带 LDL^T 分解的 DFP 对偶算法 …… 96
第7章 直接搜索方法 …… 99
7.1 单纯形替换法 …… 99
7.1.1 $\mathbf{R}^n$ 中的单纯形 …… 99
7.1.2 单纯形替换算法 …… 103
7.2 方向加速法 …… 107
7.2.1 基本定理及 Powell 基本算法 …… 107
7.2.2 Powell 算法的方向调整原理 …… 112
7.2.3 Powell 算法方向调整的判别准则 …… 114
第8章 线性最小二乘法 …… 120
8.1 观测数据的最小二乘拟合 …… 120
8.1.1 残差 …… 121
8.1.2 最小二乘拟合的数学模型 …… 122
8.2 超定方程组及其最小二乘解 …… 123
8.2.1 超定方程组的最小二乘解 …… 123
8.2.2 最小二乘解的存在性及唯一性 …… 126
8.2.3 举例 …… 129
8.3 Golub 方法(用正交分解求最小二乘解) …… 132
8.3.1 矩阵的正交分解 …… 132
8.3.2 Golub 算法 …… 135
附录 I 初等反射矩阵(H 矩阵)及其性质 …… 141

第 9 章 非线性最小二乘法 …… 146
9.1 非线性最小二乘法问题 …… 146
9.1.1 问题的提出 …… 146
9.1.2 问题的形成 …… 146
9.1.3 解法概述 …… 149
9.2 Gauss-Newton 算法(简称 G-N 算法) …… 150
9.2.1 G-N 方向的构造 …… 150
9.2.2 G-N 算法及其局部收敛性质 …… 151
9.3 修正的 G-N 算法(Hartley 方法) …… 157
9.4 Levenberg-Marquarat 算法(简称 L-M 算法) …… 159
9.4.1 L-M 算法的基本想法 …… 159
9.4.2 L-M 算法的基础定理 …… 160
9.4.3 L-M 算法 …… 165
9.4.4 L-M 算法的收敛性质 …… 166
附录 II 最优化方法的发展进程 …… 169
无约束最优化方法习题 …… 174

下篇 约束最优化方法

第 10 章 线性规划及其解法 …… 182
10.1 线性规划问题举例 …… 182
10.2 线性规划问题的基本概念及解的性质 …… 184
10.2.1 线性规划模型的一般形式 …… 184
10.2.2 线性规划问题解的概念 …… 185
10.2.3 线性规划问题解的性质 …… 187
10.3 单纯形法 …… 191
10.3.1 单纯形法原理 …… 191
10.3.2 用人工变量法找初始可行基——大 M 法和两段单纯形法 …… 195
10.3.3 修正单纯形法 …… 200
10.4 线性规划的对偶问题 …… 204
10.4.1 对偶问题举例 …… 204
10.4.2 原问题与对偶问题的关系 …… 206
10.4.3 对偶问题的基本定理 …… 208
10.4.4 对偶单纯形法 …… 212
第 11 章 整数规划 …… 215
11.1 整数规划问题举例 …… 215
11.2 整数规划的分枝定界法和割平面法 …… 217
11.2.1 分枝定界法 …… 217

11.2.2 割平面法 …… 220
11.3 0－1规划 …… 223
11.3.1 0－1规划举例 …… 223
11.3.2 0－1规划的解法 …… 224
11.4 指(分)派问题 …… 227
11.4.1 指(分)派问题举例 …… 227
11.4.2 匈牙利法 …… 228
11.5 整数规划问题应用实例 …… 231
第12章 约束最优化问题的最优性条件 …… 234
12.1 约束最优化问题的数学描述 …… 234
12.1.1 全局解与局部解 …… 234
12.1.2 凸规划 …… 235
12.2 几何最优性条件 …… 237
12.2.1 必要条件 …… 237
12.2.2 充分条件 …… 242
12.3 引用Lagrange函数的最优性条件 …… 243
12.3.1 必要条件 …… 245
12.3.2 充分条件 …… 249
第13章 非线性规划的对偶理论 …… 252
13.1 Lagrange对偶问题与弱对偶性定理 …… 252
13.2 鞍点判别条件 …… 255
13.3 扩展的对偶定理 …… 258
第14章 可行方向法 …… 261
14.1 可行方向法 …… 261
14.1.1 线性约束的情形 …… 264
14.1.2 非线性约束的情形 …… 270
14.2 投影梯度法 …… 277
14.3 既约梯度法 …… 284
第15章 罚函数法 …… 292
15.1 罚函数法 …… 292
15.1.1 罚函数法 …… 292
15.1.2 罚函数法的收敛性质 …… 295
15.2 障碍函数法 …… 297
15.2.1 算法的构成 …… 297
15.2.2 障碍函数法的收敛性定理 …… 300
15.3 广义Lagrange乘子法 …… 301
15.3.1 等式约束下的广义乘子法 …… 302
15.3.2 不等式约束下的广义乘子法 …… 305

15.3.3 等式与不等式约束下的广义乘子法 …… 306
15.4 精确罚函数法 …… 308
15.4.1 非线性等式约束问题的可微精确罚函数法 …… 308
15.4.2 一般非线性约束问题的可微精确罚函数法 …… 309
第16章 二次规划 …… 311
16.1 二次规划问题及其 k-T 条件 …… 311
16.2 Lemke 算法 …… 313
16.3 Wolfe 方法 …… 317
16.4 序列二次规划法 …… 320
第17章 离散系统的动态规划方法 …… 324
17.1 多阶段决策问题(引例及相关基本概念) …… 324
17.2 多阶段决策问题的数学描述 …… 327
17.2.1 数学模型 …… 327
17.2.2 Bellman 最优性原理 …… 327
17.2.3 动态规划基本定理 …… 328
17.3 求解多阶段决策问题的动态规划方法 …… 330
17.4 多阶段决策问题实例分析 …… 334
17.5 离散线性二次型系统的动态规划方法 …… 338
第18章 现代优化方法简介 …… 342
18.1 随机试验法 …… 342
18.2 禁忌搜索算法 …… 346
18.2.1 禁忌搜索算法的主要步骤 …… 346
18.2.2 禁忌搜索算法的特征 …… 347
18.3 模拟退火算法 …… 350
18.3.1 模拟退火算法的基本原理 …… 351
18.3.2 模拟退火算法的基本步骤和实现的技术问题 …… 352
18.4 遗传算法 …… 356
18.4.1 遗传算法的基本原理和步骤 …… 356
18.4.2 遗传算法的技术问题 …… 357
18.5 神经网络算法 …… 361
18.5.1 人工神经网络的基本概念 …… 361
18.5.2 人工神经网络的基本模型 …… 362
18.5.3 前向型人工神经网络 …… 365
18.5.4 反馈型神经网络——Hopfield 模型 …… 369
附录Ⅲ Matlab 及其应用 …… 371
1.1 Matlab 简介 …… 371
1.1.1 数学软件 …… 371
1.1.2 什么是 Matlab …… 371

1.1.3 Matlab的主要用途 …… 375
1.1.4 几点说明 …… 376
1.1.5 矩阵运算 …… 376
1.2 最优化方法计算 …… 380
1.2.1 无约束极值算例 …… 380
1.2.2 约束极值 …… 382
1.2.3 线性最小二乘问题 …… 385
1.2.4 非线性最小二乘问题 …… 387
1.3 数据分析 …… 389
1.3.1 数据的输入和输出 …… 389
1.3.2 列数据分析 …… 391
1.3.3 实测数据归一化(标准化) …… 392
1.3.4 多项式拟合 …… 393
1.3.5 多元线性回归 …… 396
约束最优化方法习题 …… 399
参考文献 …… 410

上　篇
无约束最优化方法

第1章　基础知识

本章介绍最优化方法的理论基础:凸集和凸函数的概念及其基本性质;函数极值的最优性条件;凸函数极值的最优性条件.

1.1　凸集及其基本性质

定义 1.1　非空集合 $S \subset \mathbf{R}^n$ 称为凸集,若对于任意 $x_1, x_2 \in S$ 和任意实数 $\lambda \in [0,1]$,使连接 x_1, x_2 两点的线段上的任意点 $x \in S$,即

$$x = \lambda x_1 + (1-\lambda)x_2 \in S.$$

二维情形如图 1-1 所示.

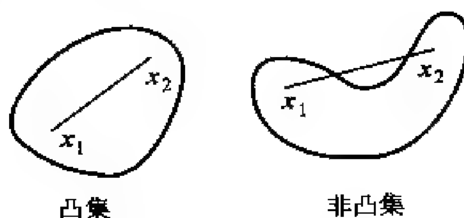


图 1-1

显然,平面上的圆域是凸集,$\mathbf{R}^n$ 中以点 $x^* \in \mathbf{R}^n$ 为心的球 $S = \{x \mid \|x - x^*\| < \delta\}$ 是凸集.

定义 1.2　设非空集合 $S \subset \mathbf{R}^n$,称集合 $H(S)$ 为 S 的凸包,若 $x \in H(S)$,当且仅当

$$x = \sum_{j=1}^{k} \lambda_j x_j,$$

其中 $\sum_{j=1}^{k} \lambda_j = 1 (\lambda_j \geqslant 0, j = 1, \cdots, k), x_j \in S (j = 1, \cdots, k)$, k 为正整数.

二维情形如图 1 - 2 所示.

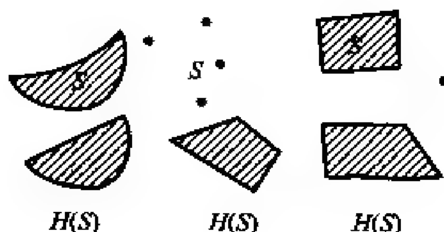


图 1 - 2

从图 1 - 2 看到,$H(S)$是所有包含 S 的最小凸集.

定义 1.3 $\mathbf{R}^n$ 中有限个点 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{k+1}$所成的凸包 $H(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{k+1})$ 称为凸多面体;如果 $\boldsymbol{x}_2-\boldsymbol{x}_1,\boldsymbol{x}_3-\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{k+1}-\boldsymbol{x}_1$ 是线性独立的,则称 $H(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{k+1})$是以 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{k+1}$为顶点的单纯形.

二维情形如图 1 - 3 所示.

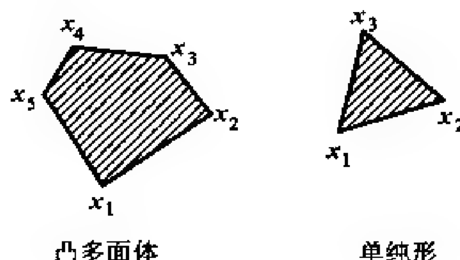


图 1 - 3

注意到在 $\mathbf{R}^n$ 中线性独立的向量的最大个数为 n,因此,在 $\mathbf{R}^n$ 中不存在多于 $n+1$ 个顶点的单纯形.

定义 1.4 非空集合 $S\subset\mathbf{R}^n$ 称为凸锥,若对于任意 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S$ 和任意实数 $\alpha\geqslant0,\beta\geqslant0$,使连线

$$\alpha\boldsymbol{x}_1+\beta\boldsymbol{x}_2\in S.$$

下面介绍凸集或与凸集有关的几个重要性质:

定理 1.1(线性性质) 设 S_1,S_2 为凸集,α 为一定数,则如下集合

$$\alpha S_1=\{\boldsymbol{y}\mid\boldsymbol{y}=\alpha\boldsymbol{x},\boldsymbol{x}\in S_1\};$$

$$S_1+S_2=\{\boldsymbol{x}\mid\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2,\boldsymbol{x}_1\in S_1,\boldsymbol{x}_2\in S_2\};$$

$$S_1\cap S_2=\{\boldsymbol{x}\mid\boldsymbol{x}\in S_1,\boldsymbol{x}\in S_2\}$$

均为凸集(读者容易验证该线性性质).

定理 1.2(投影定理) 设 S 为非空、闭集,$\boldsymbol{y}_0\notin S$,则有 $\boldsymbol{x}_0\in S$ 使

$$\|\boldsymbol{x}_0-\boldsymbol{y}_0\|=\inf_{\boldsymbol{x}\in S}\|\boldsymbol{x}-\boldsymbol{y}_0\|$$

(如图 1 - 4).

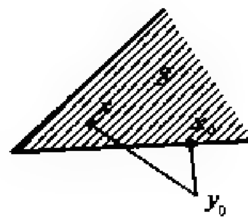


图 1 - 4

证 设 S 为有限点集，定理显然成立. 若 S 为无限点集，由下确界定义，对于 $k=1,2,\cdots$，有

$$\inf_{x\in S}\|x-y_0\| \leqslant \|x_k-y_0\| \leqslant \inf_{x\in S}\|x-y_0\| + \frac{1}{k},$$

而
$$\|x_k\| = \|x_k-y_0+y_0\| \leqslant \|x_k-y_0\| + \|y_0\|$$
$$\leqslant \inf_{x\in S}\|x-y_0\| + \frac{1}{k} + \|y_0\| \leqslant \inf_{x\in S}\|x-y_0\| + 1 + \|y_0\| \xlongequal{\text{令}} c,$$

故 $\{x_k\}$ 有界，必存在子序列 $\{x_{k_i}\}$ 使

$$\lim_{k_i\to\infty} x_{k_i} = x_0.$$

由 S 闭，于是 $x_0\in S$. 又由

$$\inf_{x\in S}\|x-y_0\| \leqslant \|x_{k_i}-y_0\| \leqslant \inf_{x\in S}\|x-y_0\| + \frac{1}{k_i},$$

令 $k_i\to\infty$ 有

$$\inf\|x-y_0\| \leqslant \|x_0-y_0\| \leqslant \inf_{x\in S}\|x-y_0\|,$$

所以
$$\|x_0-y_0\| = \inf_{x\in S}\|x-y_0\|.$$

定理 1.3（分离平面定理） 设 S 非空，且为闭凸集，$y_0\notin S$，则存在非零向量 p 及常数 α，使对任意 $x\in S$ 都有 $p^{\mathrm{T}}x\geqslant\alpha$，$p^{\mathrm{T}}y_0<\alpha$.（如图 1－5）

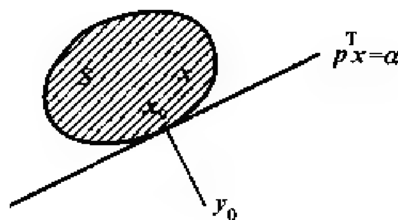


图 1－5

证 由定理 1.2 知，存在 $x_0\in S$ 使

$$\|x_0-y_0\| = \inf_{x\in S}\|x-y_0\| > 0,$$

由 S 为凸集，对任给 $\lambda\in(0,1)$ 总有

$$\lambda x+(1-\lambda)x_0\in S,\quad x\in S,$$

故
$$\|\lambda x+(1-\lambda)x_0-y_0\| \geqslant \|x_0-y_0\|,$$

即
$$\|\lambda x+(1-\lambda)x_0-y_0\|^2 \geqslant \|x_0-y_0\|^2,$$

而
$$\|\lambda x+(1-\lambda)x_0-y_0\|^2 = \|\lambda(x-x_0)+(x_0-y_0)\|^2$$
$$=\lambda^2\|x-x_0\|^2+2\lambda(x_0-y_0)^{\mathrm{T}}(x-x_0)+\|x_0-y_0\|^2,$$

因此
$$\lambda^2 \| x - x_0 \|^2 + 2\lambda (x_0 - y_0)^T (x - x_0) \geqslant 0,$$
即
$$\lambda \| x - x_0 \|^2 + 2 (x_0 - y_0)^T (x - x_0) \geqslant 0,$$
因为 $\lambda \in (0,1)$，取 λ 充分小，总有
$$2(x_0 - y_0)^T (x - x_0) \geqslant 0,$$
即
$$(x_0 - y_0)^T (x - x_0) \geqslant 0,$$
令
$$p = x_0 - y_0, \alpha = p^T x_0,$$
则有
$$p^T x \geqslant \alpha,$$
又
$$\begin{aligned} \alpha - p^T y_0 &= p^T x_0 - p^T y_0 \\ &= p^T (x_0 - y_0) \\ &= (x_0 - y_0)^T (x_0 - y_0) \\ &= \| x_0 - y_0 \|^2 > 0, \end{aligned}$$
所以
$$p^T y_0 < \alpha.$$

定理 1.4（Farkas 引理）　设 A 为给定的 $m \times n$ 实矩阵，$b \in \mathbf{R}^n$，则 $A^T y = b, y \geqslant 0$ 有解的充要条件是存在向量 x 使
$$Ax \geqslant 0,\quad b^T x \geqslant 0$$
同时满足.

证　必要性　设有 $y \geqslant 0$ 使 $A^T y = b$，则对一切满足 $Ax \geqslant 0$ 的向量 x，有
$$(Ax)^T y \geqslant 0,$$
即
$$x^T A^T y \geqslant 0,$$
从而
$$x^T b \geqslant 0.$$

充分性　设 $S = \{A^T y \mid y \geqslant 0\}$，易证 S 为闭凸集. 现证明，若 $Ax \geqslant 0$，$b^T x \geqslant 0$ 同时成立，则 $A^T y = b, y \geqslant 0$ 有解. 事实上，只需证明 $b \in S$.

用反证法，设 $b \notin S$，由定理 1.3 可知，必存在非零向量 p 及常数 α，使对任意 $\hat{y} \in S$ 都有
$$p^T \hat{y} \geqslant \alpha > p^T b, \tag{1-1}$$
易见 $0 \in S$，因而有
$$p^T b < 0.$$

下面将证明，对任意 $\hat{y} \in S$，又有 $p^T \hat{y} \geqslant 0$. (1-2)

用反证法. 事实上，若存在 $\bar{y} \in S$ 使 $p^T \bar{y} < 0$. 因为对任意 $k > 0$，总有 $k\bar{y} \in S$，特别取
$$k > \frac{p^T b}{p^T \bar{y}} (> 0)$$
时，则得到
$$p^T (k\bar{y}) < p^T b, \tag{1-3}$$

显见,(1 - 1)式与(1 - 3)式矛盾,故(1 - 2)式得证,因此有

$$p^T\hat{y} \geqslant 0 > p^Tb.$$

若取
$$y = e_i = (0,\cdots,0,\underset{\text{第}i\text{个分量}}{1},0,\cdots,0)^T,$$

显然,此时 $y \geqslant 0$. 记

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_i^T \\ \vdots \\ a_n^T \end{bmatrix}, \qquad \text{故 } A^Ty = A^Te_i = a_i \in S,$$

由(1 - 2)式知必有

$$p^Ta_i \geqslant 0, i = 1,2,\cdots,n,$$

或
$$a_i^Tp \geqslant 0, i = 1,2,\cdots,n,$$

即
$$Ap \geqslant 0.$$

但由于 $p^Tb < 0$,取 $X = p$,这便导致与充分性假设相矛盾,故必有 $b \in S$,即存在 y 满足 $A^Ty = b, y \geqslant 0$.

定理 1.5(择一定理) 设矩阵 $A_{m\times n}, B_{p\times n}$,则 $Ax > 0, Bx \geqslant 0$ 不同解的充要条件是存在 $\bar{u} \in \mathbf{R}^m, \bar{v} \in \mathbf{R}^p$($\bar{u} \geqslant 0, \bar{v} \geqslant 0$,且 $\bar{u} \neq 0$)使

$$A^Tu + B^Tv = 0.$$

证 显然,$Ax > 0, Bx \geqslant 0$ 等价于 $Ax \geqslant (\alpha,\cdots,\alpha)^T \in \mathbf{R}^m, \alpha > 0$, $Bx \geqslant 0$.

记
$$l = (1,\cdots,1)^T \in \mathbf{R}^m,$$

令
$$\bar{A} = \begin{bmatrix} A & -l \\ B & O \end{bmatrix}, \bar{b} = (0,\cdots,0,-1)^T \in \mathbf{R}^{n+1},$$

易证 $Ax \geqslant \alpha l, \alpha > 0, Bx \geqslant 0$ 等价于 $(x^T,\alpha)\bar{A}^T \geqslant 0, (x^T,\alpha)\bar{b} < 0$.

即
$$\bar{A}\begin{bmatrix} x \\ \alpha \end{bmatrix} \geqslant 0, \quad \bar{b}^T\begin{bmatrix} x \\ \alpha \end{bmatrix} < 0.$$

由 Farkas 引理,上式不同解的充要条件是

存在
$$y = \begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} \geqslant 0,$$

使
$$\bar{A}^T\begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} = \bar{b}, \bar{u} \in \mathbf{R}^m, \bar{v} \in \mathbf{R}^p,$$

即
$$\begin{bmatrix} A^T & B^T \\ -l^T & O \end{bmatrix}\begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} = \bar{b}.$$

故得
$$\begin{cases} A^T\bar{u} + B^T\bar{v} = 0, \\ -l^T\bar{u} = -1. \end{cases}$$

由 $-l^T\bar{u}=-1$ 得

$$\sum_{i=1}^{m}\bar{u}_i=1,\quad \bar{u}_i\geqslant 0,$$

即
$$\bar{u}\geqslant 0 \text{ 且 } \bar{u}\neq 0,$$

故得
$$A^T\bar{u}+B^T\bar{v}=0,$$
$$\bar{u}\geqslant 0,\bar{v}\geqslant 0 \text{ 且 } \bar{u}\neq 0.$$

综上有 $Ax>0,Bx\geqslant 0$ 不同解的充要条件是存在 $\bar{u},\bar{v}$ 使

$$A^Tu+B^Tv=0\ (\bar{u}\geqslant 0,\bar{v}\geqslant 0,\text{且 } \bar{u}\neq 0).$$

定理 1.4 和定理 1.5 是第 12 章中研究约束极值最优性条件的基础.

1.2 极值(一般函数)的最优性条件

关于一元函数极值的概念在微积分中已有叙述,本节针对多元函数的情形给出各类极小点的定义及局部极小点的判定条件(最优性条件).

1.2.1 多元函数极值概念

定义 1.5 对于任意给定的实数 $\delta>0$,满足不等式 $\|x-x_0\|<\delta$ 的 x 的集合称为点 x_0 的邻域,记为 $N(x_0,\delta)=\{x \mid \|x-x_0\|<\delta\}$.

定义 1.6 设 $f:S\subset \mathbf{R}^n\rightarrow \mathbf{R}^1$,若存在一个数 $\delta>0$,对 $\forall x\in N(x^*,\delta)\cap S$ 都有

$$f(x^*)\leqslant f(x),$$

则称 x^* 为 $f(x)$ 的局部极小点.

若对 $\forall x\in N(x^*,\delta)\cap S$,但 $x\neq x^*$ 有 $f(x^*)<f(x)$,则称 x^* 为 $f(x)$ 的严格局部极小点.

定义 1.7 设 $f:S\subset \mathbf{R}^n\rightarrow \mathbf{R}^1$,若对 $\forall x\in S$ 都有

$$f(x^*)\leqslant f(x),$$

则称 x^* 为 $f(x)$ 在 S 上的全局极小点.

若对 $\forall x\in S$,但 $x\neq x^*$ 有 $f(x^*)<f(x)$,则称 x^* 为 $f(x)$ 在 S 上的严格全局极小点.

由以上定义看到,x^* 是局部极小点,是指在邻域 $N(x^*,\delta)$ 中 $f(x)$ 在 x^* 取得最小的值;而 x^* 是全局极小点,是指在定义域 S 中 $f(x)$ 在 x^* 取得最小值. 全局极小点可能在某个局部极小点处取得,也可能在 S 的边界上取得.

完全类似地可以定义函数局部极大点和全局极大点.

还应该注意到,并非每一个实值函数都有极值,例如定义在 $\mathbf{R}^n$ 上的

非零线性函数就没有极值．因此，还需进一步判断$f(x)$在S上是否存在极值的问题，如果存在，又如何判别是极小值还是极大值．

1.2.2 梯度与 Hesse 矩阵

首先引入多元函数可微的概念，在此基础上形成的函数梯度及 Hesse 矩阵的概念是研究局部极值点的重要工具．

定义 1.8 设$f:S\subset\mathbf{R}^n\to\mathbf{R}^1$，$x_0\in S$，称$f(x)$在$x_0$处可微，若存在向量$l\in\mathbf{R}^n$，对于任意的$p\in\mathbf{R}^n$，使

$$\lim_{\|p\|\to 0}\frac{f(x_0+p)-f(x_0)-l^{\mathrm{T}}p}{\|p\|}=0 \qquad (1-4)$$

成立．

易看出，(1－4)式等价于关系式

$$f(x_0+p)=f(x_0)+l^{\mathrm{T}}p+o(\|p\|), \qquad (1-5)$$

因此，我们称关于p的线性函数$l^{\mathrm{T}}p$为函数增量$f(x_0+p)-f(x_0)$的线性主部，即$f(x)$在x_0处微分．下面的定理给出(1－4)式中L的表示式．

定理 1.6 若$f(x)$在x_0处可微，则$f(x)$在该点存在关于自变量的一阶偏导数，且有

$$l=\left(\frac{\partial f(x_0)}{\partial x_1},\cdots,\frac{\partial f(x_0)}{\partial x_n}\right)^{\mathrm{T}}. \qquad (1-6)$$

证 令

$$l=(l_1,l_2,\cdots,l_n)^{\mathrm{T}},$$

且分别取

$$p=p_ie_i, i=1,2,\cdots,n,$$

其中p_i为非零实数，$e_i=(0,\cdots,0,1,0,\cdots,0)^{\mathrm{T}}$．由于$f(x)$在$x_0$处可微，则由(1－5)式得：

$$f(x_0+p_ie_i)-f(x_0)=l_ip_i+o(\|p\|),\quad i=1,2,\cdots,n.$$

两边除以p_i，并取$p_i\to 0$的极限，得到

$$\lim_{p_i\to 0}\frac{f(x_0+p_ie_i)-f(x_0)}{p_i}=l_i, i=1,2,\cdots,n.$$

据偏导数的定义，有

$$l_i=\frac{\partial f(x_0)}{\partial x_i},\quad i=1,2,\cdots,n.$$

定义 1.9 以$f(x)$的n个偏导数为分量的n维向量称为$f(x)$在x处的梯度，记为

$$\nabla f(x)=\left(\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}\right)^{\mathrm{T}}. \qquad (1-7)$$

梯度也可以称为$f(x)$关于向量x的一阶导数，常记为 **grad** $f(x)$．

如果$f(x)$在x_0处可微,由(1-5)式及(1-6)式得到

$$f(x_0+p)=f(x_0)+\nabla f(x_0)^{\mathrm T}p+o(\|p\|), \qquad (1-8)$$

这个公式即为展开到两项的多元函数的 Taylor 公式.

以下几个特殊类型函数的梯度公式是常用的.

1. 若$f(x)=c$(常数),则$\nabla f(x)=0$.

2. $\nabla(b^{\mathrm T}x)=b$.

证 设$b=(b_1,b_2,\cdots,b_n)^{\mathrm T}$,$x=(x_1,x_2,\cdots,x_n)^{\mathrm T}$,

则
$$b^{\mathrm T}x=\sum_{i=1}^{n}b_ix_i,$$

于是
$$\frac{\partial}{\partial x_j}(b^{\mathrm T}x)=\frac{\partial}{\partial x_j}\left(\sum_{i=1}^{n}b_ix_i\right)=b_j, j=1,2,\cdots,n,$$

所以
$$\nabla(b^{\mathrm T}x)=b.$$

3. $\nabla(x^{\mathrm T}x)=2x$.

4. 设A为对称方阵,则

$$\nabla(x^{\mathrm T}Ax)=2Ax.$$

证
$$x^{\mathrm T}Ax=(x_1,x_2,\cdots,x_n)\begin{bmatrix}\sum_{j=1}^{n}a_{1j}x_j\\ \sum_{j=1}^{n}a_{2j}x_j\\ \vdots\\ \sum_{j=1}^{n}a_{nj}x_j\end{bmatrix}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_jx_i,$$

于是
$$\frac{\partial}{\partial x_i}(x^{\mathrm T}Ax)=2\sum_{j=1}^{n}a_{ij}x_j, i=1,2,\cdots,n.$$

所以
$$\nabla(x^{\mathrm T}Ax)=2\left(\sum_{j=1}^{n}a_{1j}x_j,\sum_{j=1}^{n}a_{2j}x_j,\cdots,\sum_{j=1}^{n}a_{nj}x_j\right)^{\mathrm T}=2Ax.$$

综合上述公式,对于二次函数

$$f(x)=\frac{1}{2}x^{\mathrm T}Ax-b^{\mathrm T}x+c,$$

我们有
$$\nabla f(x)=\nabla\left(\frac{1}{2}x^{\mathrm T}Ax-b^{\mathrm T}x+c\right)=Ax-b,$$

其中$A=(a_{ij})_{n\times n}$为实对称阵,$b\in\mathbf{R}^n$,$x\in\mathbf{R}^n$,c为常数.

如前所述,梯度$\nabla f(x)$是$f(x)$关于x的一阶导数,现在问:$f(x)$关于x的二阶导数是什么?

首先考虑向量值函数及其导数.

向量值函数$f:S\subset\mathbf{R}^n\to\mathbf{R}^m(m>1)$的具体形式为

$$f(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x}))^{\mathrm{T}},$$

其中 $f_i(\boldsymbol{x})$ 为给定在 $S \subset \mathbf{R}^n \to \mathbf{R}^1$ 上的实值函数.

对于向量值函数的可微性,我们引入以下定义.

定义 1.10　称向量值函数 $f: S \subset \mathbf{R}^n \to \mathbf{R}^m$ 在 $\boldsymbol{x}_0 \in S$ 点可微,如果 $f(\boldsymbol{x})$ 的所有分量 $f_i(\boldsymbol{x})$ $(i=1,2,\cdots,m)$ 在 $\boldsymbol{x}_0$ 点可微,即成立

$$\lim_{\|\boldsymbol{p}\| \to 0} \frac{f_i(\boldsymbol{x}_0 + \boldsymbol{p}) - f_i(\boldsymbol{x}_0) - \nabla f_i(\boldsymbol{x}_0)^{\mathrm{T}} \boldsymbol{p}}{\|\boldsymbol{p}\|} = 0. \qquad (1-9)$$

由于向量的极限是通过它的所有分量的极限定义的,所以(1-9)式等价于

$$\lim_{\|\boldsymbol{p}\| \to 0} \frac{f(\boldsymbol{x}_0 + \boldsymbol{p}) - f(\boldsymbol{x}_0) - Df(\boldsymbol{x}_0)\boldsymbol{p}}{\|\boldsymbol{p}\|} = 0, \qquad (1-10)$$

其中

$$Df(\boldsymbol{x}_0) = \begin{bmatrix} \dfrac{\partial f_1(\boldsymbol{x}_0)}{\partial x_1} & \dfrac{\partial f_1(\boldsymbol{x}_0)}{\partial x_2} & \cdots & \dfrac{\partial f_1(\boldsymbol{x}_0)}{\partial x_n} \\ \dfrac{\partial f_2(\boldsymbol{x}_0)}{\partial x_1} & \dfrac{\partial f_2(\boldsymbol{x}_0)}{\partial x_2} & \cdots & \dfrac{\partial f_2(\boldsymbol{x}_0)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial f_m(\boldsymbol{x}_0)}{\partial x_1} & \dfrac{\partial f_m(\boldsymbol{x}_0)}{\partial x_2} & \cdots & \dfrac{\partial f_m(\boldsymbol{x}_0)}{\partial x_n} \end{bmatrix}, \qquad (1-11)$$

矩阵 $Df(\boldsymbol{x}_0)$ 称为向量值函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}_0$ 处的导数,也称为 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}_0$ 处的 Jacobi 矩阵.

应注意到,尽管(1-10)和(1-4)的定义形式相似,然而,向量值函数的导数 $Df(\boldsymbol{x}_0)$ 已不再是向量了,而是由(1-11)定义的 $m \times n$ 阶矩阵.

现在我们利用多元函数及向量值函数的一阶导数来定义多元函数的二阶导数.

设 $f: S \subset \mathbf{R}^n \to \mathbf{R}^1$ 于 S 上可微,此时 $f(\boldsymbol{x})$ 的一阶导数定义了一个向量

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right)^{\mathrm{T}}.$$

若令 $\boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})$,则 $\boldsymbol{g}: S_0 \subset S \subset \mathbf{R}^n \to \mathbf{R}^n$ 为 n 维的向量值函数. 如果 $\boldsymbol{g}(\boldsymbol{x})$ 于 $S_0 \subset S$ 上可微,则 $f(\boldsymbol{x})$ 于 $S_0 \subset S$ 上两次可微,且 $D\boldsymbol{g}(\boldsymbol{x})$ 就是多元函数 $f(\boldsymbol{x})$ 的二阶导数,即

$$D\boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}, \qquad (1-12)$$

故多元函数 $f(\boldsymbol{x})$ 的二阶导数为 $n\times n$ 阶矩阵：常称矩阵(1-12)为 $f(\boldsymbol{x})$ 的 Hesse 矩阵，记为

$$H(\boldsymbol{x}) = Dg(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x}),$$

它是一元函数的二阶导数在多元情形下的推广.

综上所述，多元函数 $f(\boldsymbol{x})$ 的一阶导数是它的梯度 $\nabla f(\boldsymbol{x})$，二阶导数是它的 Hesse 矩阵 $\nabla^2 f(\boldsymbol{x})$，在最优化方法的讨论中这是两个常用的概念.

在微积分中已经证明过，当 $f(\boldsymbol{x})$ 的所有二阶偏导数连续时，有

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_j \partial x_i}, \quad i,j = 1,2,\cdots,n.$$

此时，$f(\boldsymbol{x})$ 的 Hesse 矩阵是对称矩阵.

类似于一元函数的情形，利用函数的梯度(一阶导数)及 Hesse 矩阵(二阶导数)可以表达多元函数的 Taylor 展开式.

定理 1.7 设 $f:S\subset\mathbf{R}^n\to\mathbf{R}^1$ 在开球

$$S_0 = \{\boldsymbol{x} \mid \|\boldsymbol{x}-\boldsymbol{x}_0\| < \delta, \delta > 0\} \subset S$$

上可微，则对满足 $\|\boldsymbol{p}\|<\delta$ 的 $\boldsymbol{p}$ 成立

$$f(\boldsymbol{x}_0+\boldsymbol{p}) = f(\boldsymbol{x}_0) + \nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} \tag{1-13}$$

其中 $\bar{\boldsymbol{x}} = \boldsymbol{x}_0 + \xi\boldsymbol{p}, \xi\in(0,1)$.

证 引入函数 $\phi(t) = f(\boldsymbol{x}_0 + t\boldsymbol{p})$，

于是 $\phi(0) = f(\boldsymbol{x}_0), \phi(1) = f(\boldsymbol{x}_0+\boldsymbol{p})$，

由 $f(\boldsymbol{x})$ 的可微性容易证得，对给定的 $\boldsymbol{p}\in S_0$，$\phi(t)$ 是 $[0,1]$ 上的可微函数，且有

$$\phi'(t) = \nabla f(\boldsymbol{x}_0 + t\boldsymbol{p})^{\mathrm{T}}\boldsymbol{p},$$

对任意 $t\in[0,1]$ 由一维中值定理知，存在 $\xi\in(0,1)$，成立 $\phi(1)-\phi(0) = \phi'(\xi)$，即

$$f(\boldsymbol{x}_0+\boldsymbol{p}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0+\xi\boldsymbol{p})^{\mathrm{T}}\boldsymbol{p} = f(\boldsymbol{x}_0) + \nabla f(\bar{\boldsymbol{x}})\boldsymbol{p}.$$

式(1-13)还可写成

$$f(\boldsymbol{x}_0+\boldsymbol{p}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^{\mathrm{T}}\boldsymbol{p} + o(\|\boldsymbol{p}\|). \tag{1-14}$$

其中 $o(\|\boldsymbol{p}\|) = \|\boldsymbol{p}\|\alpha(\boldsymbol{x}_0,\boldsymbol{p})$. 当 $\|\boldsymbol{p}\|\to 0$ 时，$\alpha(\boldsymbol{x}_0,\boldsymbol{p})\to 0$，请读者自证.

定理 1.8 设 $f:S\subset\mathbf{R}^n\to\mathbf{R}^1$ 在开球

$$S_0 = \{\boldsymbol{x} \mid \|\boldsymbol{x}-\boldsymbol{x}_0\| < \delta, \delta > 0\} \subset S$$

上二次可微，则对满足 $\|\boldsymbol{p}\|<\delta$ 的 $\boldsymbol{p}$ 成立

$$f(\boldsymbol{x}_0+\boldsymbol{p}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^{\mathrm{T}}\boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^{\mathrm{T}}H(\bar{\boldsymbol{x}})\boldsymbol{p}, \tag{1-15}$$

其中 $H(\bar{x}) = \nabla^2 f(\bar{x}), \bar{x} = x_0 + \xi p$，而 $\xi \in (0,1)$.

证 令 $\phi(t) = f(x_0 + tp)$，由定理假定，$\phi(t)$ 在 $[0,1]$ 上二次可微，并有

$$\phi'(t) = \nabla f(x_0 + tp)^{\mathrm{T}} p, \tag{1-16}$$

$$\phi''(t) = p^{\mathrm{T}} H(x_0 + tp) p, \tag{1-17}$$

按一元函数 Taylor 展开定理把 $\phi(t)$ 在 $t=0$ 点展开，得到：

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{1}{2}\phi''(\xi t)t^2,$$

即
$$f(x_0 + tp) = f(x_0) + \nabla f(x_0)^{\mathrm{T}} p + \frac{t^2}{2} p^{\mathrm{T}} H(x_0 + \xi t p) p.$$

令 $t=1$，得

$$f(x_0 + p) = f(x_0) + \nabla f(x_0)^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} H(x_0 + \xi p) p$$
$$= f(x_0) + \nabla f(x_0)^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} H(\bar{x}) p.$$

(1-15)式还写成

$$f(x_0 + P) = f(x_0) + \nabla f(x_0)^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} H(x_0) p + o(\| p \|^2), \tag{1-18}$$

其中 $o(\| p \|^2) = \| p \|^2 \alpha(x_0, p)$，当 $\| p \| \to 0$ 时，$\alpha(x_0, p) \to 0$，请读者自证.

1.2.3 局部极值的最优性条件

定义 1.11（下降方向） 设 $f: S \subset \mathbf{R}^n \to \mathbf{R}^1$ 在点 $\hat{x} \in S$ 处可微，$p \in \mathbf{R}^n$ 是已知的非零向量，如果

$$\nabla f(\hat{x})^{\mathrm{T}} p < 0, \tag{1-19}$$

则称 p 是 $f(x)$ 在 $\hat{x}$ 处的下降方向.

为了图示下降方向，我们以二维情形为例.

设 $f(x) = f(x_1, x_2)$ 在 $\hat{x} = (\hat{x}_1, \hat{x}_2)$ 处的等值线（图 1-6）为

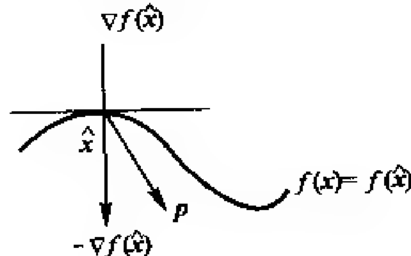


图 1-6

$$f(x_1,x_2) = f(\hat{x}_1,\hat{x}_2)$$

按定义，若 $\boldsymbol{p}$ 为下降方向，则

$$\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} = \|\nabla f(\hat{\boldsymbol{x}})\| \times \|\boldsymbol{p}\| \cos\theta < 0,$$

故 $\theta > \frac{\pi}{2}$，即下降方向 $\boldsymbol{p}$ 是靠向 $-\nabla f(\hat{\boldsymbol{x}})$ 一侧的.

由定义 1.11 可以直接验证如下事实：$\boldsymbol{p} = -\nabla f(\boldsymbol{x})$ 及 $\boldsymbol{p} = -\boldsymbol{A}\nabla f(\boldsymbol{x})$（$\boldsymbol{A}$ 为正定矩阵）都是 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的下降方向.

对于负梯度方向 $\boldsymbol{p} = -\nabla f(\boldsymbol{x})$，它不仅是下降方向，而且是最速下降方向，据方向导数的定义，设 $\nabla f(\boldsymbol{x}) \neq 0$，有

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{p}} = \nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{e} = \|\nabla f(\boldsymbol{x})\| \cos(\nabla f(\boldsymbol{x}),\boldsymbol{e}).$$

为使 $\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{p}}$ 取得最小值，只须令 $(\nabla f(\boldsymbol{x}),\boldsymbol{e}) = (\nabla f(\boldsymbol{x}),\boldsymbol{p}) = 180°$（$\boldsymbol{e}$ 是 $\boldsymbol{p}$ 方向上的单位向量），即取 $\boldsymbol{p} = -\nabla f(\boldsymbol{x})$. 因此，$f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的负梯度方向就是 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的最速下降方向.

下面进一步讨论下降方向的性质.

定理 1.9　设 $f: S \subset \mathbf{R}^n \to \mathbf{R}^1$ 在 $\hat{\boldsymbol{x}} \in S$ 处可微，$\boldsymbol{p}$ 为 $f(\boldsymbol{x})$ 在 $\hat{\boldsymbol{x}}$ 处的下降方向，则存在 $\delta > 0$，使得对 $\forall \lambda \in (0,\delta)$ 有

$$f(\hat{\boldsymbol{x}} + \lambda\boldsymbol{p}) < f(\hat{\boldsymbol{x}}). \tag{1-20}$$

证　由于 $f(\boldsymbol{x})$ 在 $\hat{\boldsymbol{x}}$ 处可微，据定理 1.7 中的(1-14)式可得

$$f(\hat{\boldsymbol{x}} + \lambda\boldsymbol{p}) = f(\hat{\boldsymbol{x}}) + \lambda\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} + \lambda\|\boldsymbol{p}\|\alpha(\hat{\boldsymbol{x}},\lambda\boldsymbol{p}),$$

其中当 $\lambda \to 0$ 时，$\alpha(\hat{\boldsymbol{x}},\lambda\boldsymbol{p}) \to 0$，由上式可得

$$\frac{f(\hat{\boldsymbol{x}} + \lambda\boldsymbol{p}) - f(\hat{\boldsymbol{x}})}{\lambda} = \nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} + \|\boldsymbol{p}\|\alpha(\hat{\boldsymbol{x}},\lambda\boldsymbol{p}),$$

由于 $\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} < 0$. 而 $\alpha(\hat{\boldsymbol{x}},\lambda\boldsymbol{p}) \to 0\ (\lambda \to 0)$，故存在一个 δ，当 $\lambda \in (0,\delta)$ 时，上式右端小于零，从而证明了(1-20)成立.

定理 1.9 表明，从 $\hat{\boldsymbol{x}}$ 出发，当 $\boldsymbol{x}$ 沿下降方向 $\boldsymbol{p}$ 移动时，$f(\boldsymbol{x})$ 的值在减小，但并非无限制地减小，而只是当 $\boldsymbol{x}$ 在 $\hat{\boldsymbol{x}}$ 的邻近时，才能保证 $f(\boldsymbol{x})$ 的值减小，这说明函数下降的局部性质.

定理 1.10　设(1)若 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处可微，且 $\boldsymbol{x}^*$ 为 $f(\boldsymbol{x})$ 的局部极小，则

$$\nabla f(\boldsymbol{x}^*) = 0; \tag{1-21}$$

(2) 若 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处二次可微，且 $\boldsymbol{x}^*$ 为 $f(\boldsymbol{x})$ 的局部极小，则 Hesse 矩阵 $H(\boldsymbol{x}^*)$ 半正定.

证　(1) 设 $\nabla f(\boldsymbol{x}^*) \neq 0$；令 $\boldsymbol{p} = -\nabla f(\boldsymbol{x}^*)$ 则 $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} = -\|\nabla f(\boldsymbol{x}^*)\|^2 < 0$，即 $\boldsymbol{p}$ 为 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处的下降方向，由(1-20)得

$$f(\boldsymbol{x}^* + \lambda\boldsymbol{p}) < f(\boldsymbol{x}^*),\ \forall \lambda \in (0,\delta),$$

这与 x^* 是 $f(x)$ 的局部极小点相矛盾，故

$$\nabla f(x^*) = 0;$$

（2）考虑任意非零向量 p，由 $f(x)$ 在 x^* 处二次可微，按式(1－18)有

$$f(x^* + \lambda p) = f(x^*) + \lambda \nabla f(x^*)^{\mathrm{T}} p + \frac{1}{2}\lambda^2 p^{\mathrm{T}} H(x^*) p + \lambda^2 \|p\|^2 \alpha(x^*, \lambda p),$$

其中当 $\lambda \to 0$ 时，$\alpha(x^*, \lambda p) \to 0$，因为 x^* 是局部极小点，故

$$\nabla f(x^*) = 0,$$

于是上式可化为

$$\frac{f(x^* + \lambda p) - f(x^*)}{\lambda^2} = \frac{1}{2} p^{\mathrm{T}} H(x^*) p + \|p\|^2 \alpha(x^*, \lambda p).$$

又因为 x^* 是局部极小，故对充分小的 λ

$$f(x^* + \lambda p) \geqslant f(x^*),$$

由上式可知，对充分小的 λ 有

$$\frac{1}{2} p^{\mathrm{T}} H(x^*) p + \|p\|^2 \alpha(x^*, \lambda p) \geqslant 0,$$

令 $\lambda \to 0$ 得知 $$p^{\mathrm{T}} H(x^*) p \geqslant 0,$$

因此，$H(x^*)$ 是半正定的.

条件(1－21)仅仅是必要的，而不是充分的. 例如 $f(x_1, x_2) = x_1 x_2$ 在点 $x^* = (0,0)^{\mathrm{T}}$ 处的梯度是 $\nabla f(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. 但是点 x^* 是双曲面的鞍点，而不是极小点（图 1－7）.

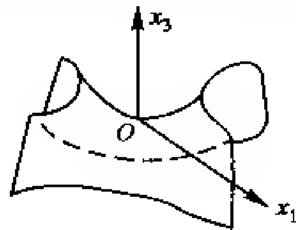


图 1－7

今后，我们将满足(1－21)的点统称为稳定点或驻点. 在比较简单的情况下，为了求得函数的极小点，可先通过(1－21)将所有的稳定点求出来，然后进一步判断其中哪些是极小点，然而在一般情况下，上述作法是很难实现的. 因此有必要引入判断极小点的充分条件.

定理 1.11（充分最优性条件）　设 $f: S \subset \mathbf{R}^n \to \mathbf{R}^1$，在 $x^* \in S$ 处二次连续可微，如果 $\nabla f(x^*) = 0$ 且 $H(x^*)$ 正定，则 x^* 是 $f(x)$ 的严格局部极小点.

证　欲证存在 x^* 的一个邻域 $N(x^*, \delta) \subset S$，使得 $\forall x \in N(x^*, \delta)$，但 $x \neq x^*$ 都有

$$f(x) > f(x^*).$$

现令 $x = x^* + tp$，其中 p 为单位向量. 引入辅助函数

$$\varphi(t) = f(x^* + tp),$$

在 $t = 0$ 点展开成 Taylor 表达式

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{1}{2}\phi''(\theta t)t^2, \qquad (1-22)$$

其中 $0<\theta<1$. 由假设

$$\varphi'(0) = \nabla f(x^*)^{\mathrm{T}} p = 0,$$

因此(1－22)变成

$$\phi(t) - \phi(0) = \frac{1}{2}\phi''(\theta t)t^2. \qquad (1-23)$$

再由假设

$$\phi''(0) = p^{\mathrm{T}} \nabla^2 f(x^*)p = p^{\mathrm{T}} H(x^*)p > 0,$$

又根据 $\varphi''(t)$ 的连续性,存在 $\varepsilon>0$,对一切 $\theta\in(0,1)$ 和满足 $|t|<\varepsilon$ 的 t,成立

$$\phi''(\theta t) > 0. \qquad (1-24)$$

联立(1－23)、(1－24)立刻得结论

$$f(x) > f(x^*).$$

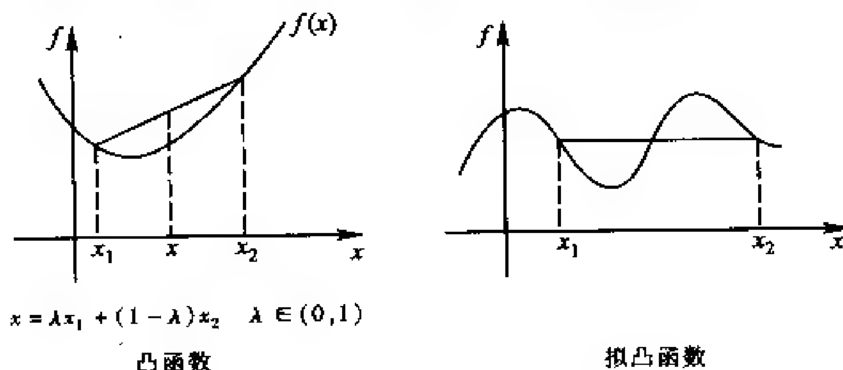
1.3　凸函数及凸函数极值的最优性条件

凸函数有许多重要的性质,例如,凸函数在凸集内部上的任何局部极小也是整体极小. 下面就给出凸函数的概念及判定,并阐述凸函数的某些性质,进而给出凸函数的最优性条件.

1.3.1　凸函数的定义及判定

定义 1.12　设 $f: S\subset \mathbf{R}^n \to \mathbf{R}^1$, S 是一个凸集,若对 $\forall x_1, x_2 \in S, \lambda \in (0,1)$,有 $f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda)f(x_2)$,则称 $f(x)$ 是 S 上的凸函数. 若上式中不等号严格成立,则称 $f(x)$ 是严格凸函数.

凸函数的几何意义:对于一元函数 $y=f(x)$ 的情形,如图 1－8 所示:



凸函数　　　　拟凸函数

图 1－8

定义 1.13（上图，下图）　设 S 是一凸集，且非空，α 为任意实数，f: $S\subset \mathbf{R}^n\to\mathbf{R}^1$，对 $\mathbf{R}^{n+1}$ 中的集合：

$$\mathbf{epi}\, f=\{(x,\alpha)\mid x\in S, f(x)\leqslant\alpha\}\ \text{称为}\ f(x)\ \text{的上图},$$

$$\mathbf{hyp}\, f=\{(x,\alpha)\mid x\in S, f(x)\geqslant\alpha\}\ \text{称为}\ f(x)\ \text{的下图}.$$

上图、下图的几何意义：对于一元函数 $f:S\subset\mathbf{R}^n\to\mathbf{R}^1$（$S$ 为凸函数）的情形，如图 1－9 所示：

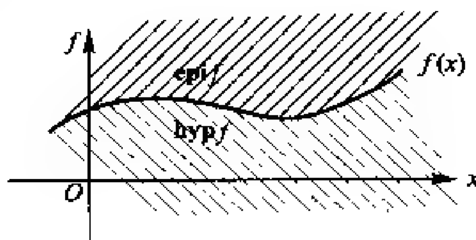


图 1－9

判定一个函数是否为凸函数，可以直接利用定义 1.12，也可以利用下面介绍的几个定理（定理 1.12—定理 1.15）.

定理 1.12　设 $S\subset\mathbf{R}^n$ 为凸集，$f:S\to\mathbf{R}^1$，则 $f(x)$ 是 S 上的凸函数的充分必要条件是 $\mathbf{epi}\, f$ 是凸集.

证　必要性　对 $\forall(x_1,\alpha_1),(x_2,\alpha_2)\in\mathbf{epi}\, f$，有

$$x_1,x_2\in S\ \text{且}\ f(x_1)\leqslant\alpha_1,\quad f(x_2)\leqslant\alpha_2,$$

从而对 $\forall\lambda\in(0,1)$，由 S 的凸性知，$\lambda x_1+(1-\lambda)x_2\in S$，

再由 $f(x)$ 是 S 上的凸函数，有

$$f(\lambda x_1+(1-\lambda)x_2)\leqslant\lambda f(x_1)+(1-\lambda)f(x_2)\leqslant\lambda\alpha_1+(1-\lambda)\alpha_2,$$

又
$$\begin{aligned}&\lambda(x_1,\alpha_1)+(1-\lambda)(x_2,\alpha_2)\\&=(\lambda x_1+(1-\lambda)x_2,\lambda\alpha_1+(1-\lambda)\alpha_2),\end{aligned}$$

故　$$\lambda(x_1,\alpha_1)+(1-\lambda)(x_2,\alpha_2)\in\mathbf{epi}\, f\quad(\text{对}\ \forall\lambda\in(0,1)),$$

即 $\mathbf{epi}\, f$ 是凸集.

充分性　对 $\forall x_1,x_2\in S,\lambda\in(0,1)$，由 S 的凸性知，$\lambda x_1+(1-\lambda)x_2\in S$，

又
$$\mathbf{epi}\, f=\{(x,\alpha)\mid x\in S,f(x)\leqslant\alpha\},$$

故
$$(x_1,f(x_1)),(x_2,f(x_2))\in\mathbf{epi}\, f.$$

由 $\mathbf{epi}\, f$ 是凸集，知

$$\begin{aligned}&\lambda(x_1,f(x_1))+(1-\lambda)(x_2,f(x_2))\\&=(\lambda x_1+(1-\lambda)x_2,\lambda f(x_1)+(1-\lambda)f(x_2))\in\mathbf{epi}\, f,\end{aligned}$$

从而
$$f(\lambda x_1+(1-\lambda)x_2)\leqslant\lambda f(x_1)+(1-\lambda)f(x_2),$$

故 $f(x)$ 是 S 上的凸函数.

定理 1.13　设 $S\subset \mathbf{R}^n$ 为凸集，$f:S\subset \mathbf{R}^n\to \mathbf{R}^1$ 在内部连续可微. 则 $f(\boldsymbol{x})$ 是内部上的凸函数的充分必要条件是对 $\forall \bar{\boldsymbol{x}},\boldsymbol{x}\in \text{int } S$，有 $f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})+\nabla^{\mathrm{T}}f(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})$.

证　必要性　由题设，对一切 $\bar{\boldsymbol{x}},\boldsymbol{x}\in \text{int } S,\lambda\in(0,1)$

有
$$f(\lambda\boldsymbol{x}+(1-\lambda)\bar{\boldsymbol{x}})\leqslant \lambda f(\boldsymbol{x})+(1-\lambda)f(\bar{\boldsymbol{x}}),$$

于是
$$\frac{f(\bar{\boldsymbol{x}}+\lambda(\boldsymbol{x}-\bar{\boldsymbol{x}}))-f(\bar{\boldsymbol{x}})}{\lambda}\leqslant f(\boldsymbol{x})-f(\bar{\boldsymbol{x}}),$$

亦即　$\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})+\lambda\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|\alpha(\bar{\boldsymbol{x}},\lambda(\boldsymbol{x}-\bar{\boldsymbol{x}}))\leqslant f(\boldsymbol{x})-f(\bar{\boldsymbol{x}})$,

上式中，令 $\lambda\to 0$，则有　$\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})\leqslant f(\boldsymbol{x})-f(\bar{\boldsymbol{x}})$,

即
$$f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})+\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}}).$$

充分性　设对 $\forall \bar{\boldsymbol{x}},\boldsymbol{x}\in \text{int } S, f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})+\nabla^{\mathrm{T}}f(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})$ 成立.

令 $z=\lambda\bar{\boldsymbol{x}}+(1-\lambda)\boldsymbol{x},\lambda\in(0,1)$，则有
$$f(\bar{\boldsymbol{x}})\geqslant f(z)+\nabla f(z)^{\mathrm{T}}(\bar{\boldsymbol{x}}-z),$$
$$f(\boldsymbol{x})\geqslant f(z)+\nabla f(z)^{\mathrm{T}}(\boldsymbol{x}-z),$$

于是 $\lambda f(\bar{\boldsymbol{x}})+(1-\lambda)f(\boldsymbol{x})\geqslant f(z)+\nabla f(z)^{\mathrm{T}}(\lambda\bar{\boldsymbol{x}}-(1-\lambda)\boldsymbol{x}-z)$
$$=f(z)=f(\lambda\bar{\boldsymbol{x}}+(1-\lambda)\boldsymbol{x}).$$

由定义 1.10 知，$f(\boldsymbol{x})$ 是 int S 上的凸函数.

定理 1.13 的几何特征如图 1－10 所示：

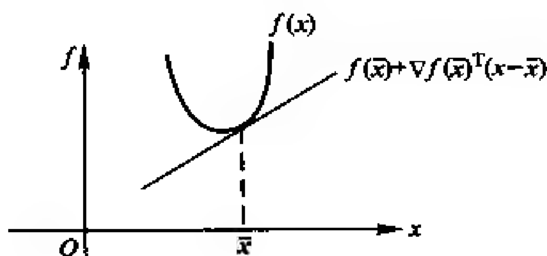


图 1－10

定理 1.14　设 $S\subset \mathbf{R}^n$ 为凸集，$f:S\subset \mathbf{R}^n\to \mathbf{R}^1$ 在 int S 二次连续可微. 则 $f(\boldsymbol{x})$ 是 int S 上的凸函数的必要充分条件是对 $\forall \boldsymbol{x}\in \text{int } S$，$\nabla^2 f(\boldsymbol{x})\geqslant 0$.

证　因 $f:S\to \mathbf{R}^1$ 在 int S 二次连续可微，故对 $\forall \bar{\boldsymbol{x}},\boldsymbol{x}\in \text{int } S$，有
$$f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})+\nabla f(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})+\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}}\nabla^2 f(\bar{\boldsymbol{x}}+\xi(\boldsymbol{x}-\bar{\boldsymbol{x}}))(\boldsymbol{x}-\bar{\boldsymbol{x}}).$$
$$(\text{其中 }\xi\in(0,1))\qquad (1-25)$$

必要性　（反证法）若存在 $\bar{\boldsymbol{x}}\in \text{int } S$，使得 $\nabla^2 f(\boldsymbol{x})<0$，即对 $\boldsymbol{x}\in \text{int } S$，$\boldsymbol{x}\neq\bar{\boldsymbol{x}}$，总有

$$(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{H}(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})<0 \qquad (\boldsymbol{H}(\bar{\boldsymbol{x}})=\nabla^2 f(\bar{\boldsymbol{x}})),$$

由 $\boldsymbol{H}(\boldsymbol{x})$ 的连续性，总可找到 $\boldsymbol{x}\in \mathrm{int}\, S, \boldsymbol{x}\neq\bar{\boldsymbol{x}}$，使得

$$(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}}\nabla^2 f(\bar{\boldsymbol{x}}+\xi(\boldsymbol{x}-\bar{\boldsymbol{x}}))(\boldsymbol{x}-\bar{\boldsymbol{x}})<0.$$

由(1-25)式，可得 $f(\boldsymbol{x})<f(\bar{\boldsymbol{x}})+\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})$，

这与 $f(\boldsymbol{x})$ 为 int S 上的凸函数相矛盾(由定理 1.13).

充分性　对 $\forall\bar{\boldsymbol{x}},\boldsymbol{x}\in \mathrm{int}\, S, \xi\in(0,1)$，由定义 1.10

$$\xi\boldsymbol{x}+(1-\xi)\bar{\boldsymbol{x}}=\bar{\boldsymbol{x}}+\xi(\boldsymbol{x}-\bar{\boldsymbol{x}})\in \mathrm{int}\, S.$$

由题设，$\nabla^2 f(\bar{\boldsymbol{x}}+\xi(\boldsymbol{x}-\bar{\boldsymbol{x}}))\geqslant 0$ 代入(1-25)式，就有

$$f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})+\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}}),$$

由定理 1.13 知 $f(\boldsymbol{x})$ 为 int S 上的凸函数.

从充分性的证明过程可以看出，若 $\nabla^2 f(\boldsymbol{x})$ 对一切 $\boldsymbol{x}\in \mathrm{int}\, S$ 都是正定的，则 $f(\boldsymbol{x})$ 在 S 的内部严格凸，但反过来不一定成立.

1.3.2　凸函数的次梯度

下面我们引进凸函数的次梯度，并阐明在凸集内部，凸函数与次梯度的关系(定理 1.15)以及次梯度在可微条件下的存在唯一性(定理 1.16).

定义 1.14　设 $S\subset\mathbf{R}^n$，非空凸，$f:S\to\mathbf{R}^1$ 是 S 上的凸函数，若 $\exists\boldsymbol{g}\in\mathbf{R}^n$，使得 $f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})+\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})$ 对 $\forall\boldsymbol{x}\in S$ 均成立，则称 $\boldsymbol{g}$ 是 $f(\boldsymbol{x})$ 在 $\bar{\boldsymbol{x}}$ 处的次梯度.

次梯度的几何解释在一维情形下，如图 1-11 所示：

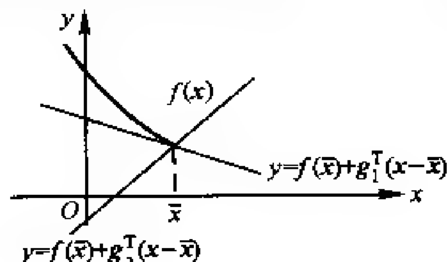


图 1-11

定理 1.15　设 S 是 $\mathbf{R}^n$ 上的凸集，$f:S\to\mathbf{R}^1$，

(1) 若 $f(\boldsymbol{x})$ 是 S 上的凸函数，则对 $\forall\bar{\boldsymbol{x}}\in \mathrm{int}\, S$，存在 f 在 $\bar{\boldsymbol{x}}$ 处的次梯度，即 $f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})+\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})$ 对 $\forall\boldsymbol{x}\in S$ 均成立；

(2) 若对 $\forall\bar{\boldsymbol{x}}\in \mathrm{int}\, S$，$\exists\boldsymbol{g}\in\mathbf{R}^n$，使得 $f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})+\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})$ 对 $\forall\boldsymbol{x}\in S$ 均成立，则 $f(\boldsymbol{x})$ 是 int S 上的凸函数；

(3) $f(\boldsymbol{x})$ 是 int S 上的凸函数的必要充分条件是对 $\forall\bar{\boldsymbol{x}}\in \mathrm{int}\, S$，存在次梯度 $\boldsymbol{g}\in\mathbf{R}^n$，满足 $f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})+\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})$，对 $\forall\boldsymbol{x}\in S$ 均成立.

证　(1) 由定理 1.12，epi f 是凸的，又 $(\bar{x}, f(\bar{x}))$ 属于 epi f 的边界，根据凸集的每一边界上有一支撑超平面知，存在一个非零向量 $(g_0, \mu) \in \mathbf{R}^n \times \mathbf{R}^1$，使得

$$g_0^T(x - \bar{x}) + \mu(\alpha - f(\bar{x})) \leqslant 0, \qquad (1-26)$$

对一切 $(x, \alpha) \in$ epi f 成立.

显然 μ 非正，否则通过选取 α 充分大，上述不等式将产生矛盾，下面我们进一步证明

$$\mu < 0.$$

假定 $\mu < 0$ 不成立，μ 必等于零，则对一切 $x \in S$，有 $g_0^T(x - \bar{x}) \leqslant 0$.

由于 $\bar{x} \in \operatorname{int} S$，存在一个 $\lambda > 0$ 使得 $\bar{x} + \lambda g_0 \in S$，从而 $\lambda g_0^T g_0 \leqslant 0$，这就意味着 $g_0 = 0$ 和 $(g_0, \mu) = (0, 0)$，这与 (g_0, μ) 是一个非零向量的事实相矛盾，因此 $\mu < 0$. 用 $|\mu|$ 除不等式(1-26)，并用 g 表示 $\frac{g_0}{|\mu|}$，就得到

$$g^T(x - \bar{x}) - \alpha + f(\bar{x}) \leqslant 0, \qquad (1-27)$$

对一切 $(x, \alpha) \in$ epi f 成立.

特别地，对一切 $x \in S$，$(x, f(x)) \in$ epi f，在(1-27)中令 $\alpha = f(x)$，就得到

$$f(x) \geqslant f(\bar{x}) + g^T(x - \bar{x}),$$

亦即，对 $\forall \bar{x} \in \operatorname{int} S$，存在 f 在 $\bar{x}$ 处的次梯度. 证明完成.

(2) 设 $\bar{x}, x \in \operatorname{int} S$，$\lambda \in (0,1)$，由 S 的凸性，易知 int S 是凸的，从而有 $\lambda x_1 + (1-\lambda)x_2 \in \operatorname{int} S$. 由假定 f 在 $\lambda x_1 + (1-\lambda)x_2$ 存在有次梯度 g. 特别地，下面的两个不等式成立：

$$f(x_1) \geqslant f(\lambda x_1 + (1-\lambda)x_2) + (1-\lambda)g^T(x_1 - x_2),$$

$$f(x_2) \geqslant f(\lambda x_1 + (1-\lambda)x_2) + \lambda g^T(x_2 - x_1).$$

用 λ 和 $(1-\lambda)$ 分别乘以上述两个不等式，并相加就得到

$$\lambda f(x_1) + (1-\lambda)f(x_2) \geqslant f(\lambda x_1 + (1-\lambda)x_2),$$

故 $f(x)$ 在 int S 上是凸的.

(3) 显然由(1)和(2)可得(3).

定理 1.16(次梯度的存在唯一性)　设 $S \subset \mathbf{R}^n$ 为凸集，$f: S \to \mathbf{R}^1$ 是凸函数，且 $f(x)$ 在 $\bar{x} \in \operatorname{int} S$ 可微，则 $f(x)$ 在 $\bar{x}$ 处的次梯度唯一，即为 $\nabla f(\bar{x})$.

证　设 g 是 $f(x)$ 在 $\bar{x}$ 处的一个次梯度，故对 $\forall d$ 及充分小的 $\lambda > 0$，有

$$f(\bar{x} + \lambda d) \geqslant f(\bar{x}) + \lambda g^T d, \qquad (1-28)$$

由 f 在 $\bar{x}$ 处可微，有

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla^T f(\bar{x}) d + o(\|\lambda d\|), \qquad (1-29)$$

其中，$o(\|\lambda \boldsymbol{d}\|)=\lambda\|\boldsymbol{d}\|\alpha(\bar{\boldsymbol{x}},\lambda \boldsymbol{d})$，$\lambda\to 0$ 时，$\alpha\to 0$.

(1－28)和(1－29)相减，得

$$0\geqslant \lambda(\boldsymbol{g}-\nabla f(\bar{\boldsymbol{x}}))^{\mathrm{T}}\boldsymbol{d}-\lambda\|\boldsymbol{d}\|\alpha(\bar{\boldsymbol{x}},\lambda \boldsymbol{d}).$$

因 $\lambda>0$，从而上式等价于

$$(\boldsymbol{g}-\nabla f(\bar{\boldsymbol{x}}))^{\mathrm{T}}\boldsymbol{d}-\|\boldsymbol{d}\|\alpha(\bar{\boldsymbol{x}},\lambda \boldsymbol{d})\leqslant 0,$$

令 $\lambda\to 0^{+}$，则对 $\forall \boldsymbol{d}$，$(\boldsymbol{g}-\nabla f(\bar{\boldsymbol{x}}))^{\mathrm{T}}\boldsymbol{d}\leqslant 0$.

特取 $\boldsymbol{d}=\boldsymbol{g}-\nabla f(\bar{\boldsymbol{x}})$，则有 $\|\boldsymbol{g}-\nabla f(\bar{\boldsymbol{x}})\|_2^2\leqslant 0$，显然，该不等式成立的充分必要条件为 $\boldsymbol{g}=\nabla f(\bar{\boldsymbol{x}})$，从而 $f(\boldsymbol{x})$ 在 $\bar{\boldsymbol{x}}$ 处的次梯度唯一，且为 $\nabla f(\bar{\boldsymbol{x}})$.

1.3.3　凸函数的最优性条件

这一部分，我们讨论凸函数在凸集上的极小问题，给出其局部极小与全局极小的关系，进而阐明凸函数的全局最优性条件.

首先，我们给出局部极小点与全局极小点的关系.

定理 1.17　设 $S\subset \mathbf{R}^n$ 为凸集，$f:S\to \mathbf{R}^1$，假定 $\bar{\boldsymbol{x}}\in S$ 是 $f(\boldsymbol{x})$ 在 S 上的一个局部极小点，

(1) 如果 f 是凸的，则 $\bar{\boldsymbol{x}}$ 是 f 在 S 上的全局极小点；

(2) 如果 f 是严格凸的，则 $\bar{\boldsymbol{x}}$ 是 f 在 S 上的唯一的全局极小点.

证　首先证明(1).

假定 f 是凸的，因为 $\bar{\boldsymbol{x}}$ 是一个局部极小点，则在 $\bar{\boldsymbol{x}}$ 附近存在一个 ε 邻域 $N_\varepsilon(\bar{\boldsymbol{x}})$，使得

$$f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}}),\qquad (1-30)$$

对 $\forall \boldsymbol{x}\in S\cap N_\varepsilon(\bar{\boldsymbol{x}})$ 成立.

现假定 $\bar{\boldsymbol{x}}$ 不是全局极小点，因而存在 $\hat{\boldsymbol{x}}\in S$，$f(\hat{\boldsymbol{x}})<f(\bar{\boldsymbol{x}})$.

由 f 的凸性，对每一个 $\lambda\in(0,1)$，下式成立：

$$f(\lambda\hat{\boldsymbol{x}}+(1-\lambda)\bar{\boldsymbol{x}})\leqslant\lambda f(\hat{\boldsymbol{x}})+(1-\lambda)f(\bar{\boldsymbol{x}})<\lambda f(\bar{\boldsymbol{x}})+(1-\lambda)f(\bar{\boldsymbol{x}})=f(\bar{\boldsymbol{x}}),$$

但当 $\lambda>0$ 且充分小时，$\lambda\hat{\boldsymbol{x}}+(1-\lambda)\bar{\boldsymbol{x}}\in S\cap \mathbf{N}_\varepsilon(\bar{\boldsymbol{x}})$，因此上述不等式与(1－30)矛盾，从而(1)得证.

(2) 现在假定 f 是严格凸的，因为严格凸性隐含凸性，则由(1)知，$\bar{\boldsymbol{x}}$ 是 f 在 S 上的全局极小点. 由反证法，假设 $\bar{\boldsymbol{x}}$ 不是唯一的全局极小点，则 $\exists \boldsymbol{x}\in S$，$\boldsymbol{x}\neq\bar{\boldsymbol{x}}$，使得 $f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})$. 由 f 的严格凸性，就有

$$f\left(\frac{1}{2}\boldsymbol{x}+\frac{1}{2}\bar{\boldsymbol{x}}\right)<\frac{1}{2}f(\boldsymbol{x})+\frac{1}{2}f(\bar{\boldsymbol{x}})=f(\bar{\boldsymbol{x}}).$$

由 S 的凸性，$\frac{1}{2}\boldsymbol{x}+\frac{1}{2}\bar{\boldsymbol{x}}\in S$，于是上述不等式违背了 $\bar{\boldsymbol{x}}$ 的全局极小性. 从

而(2)得证.

下面,我们研究 $\bar{x}$ 是全局极小点的必要充分条件.

定理 1.18 设 $S\subset \mathbf{R}^n$,非空凸,$f:S\to\mathbf{R}^1$ 是凸函数,则 $\bar{x}\in S$ 为 f 的整体(即全局)极小点的必要充分条件是 $f(x)$ 在 $\bar{x}$ 处存在一个次梯度 g,使得对一切的 $x\in S$,$g^{\mathrm{T}}(x-\bar{x})\geqslant 0$.

证 充分性 假定对一切的 $x\in S$,$g^{\mathrm{T}}(x-\bar{x})\geqslant 0$,其中 g 是 $f(x)$ 在 $\bar{x}$ 处的次梯度,由 f 的凸性(定义 1.12),有 $f(x)\geqslant f(\bar{x})+g^{\mathrm{T}}(x-\bar{x})\geqslant f(\bar{x})$,对 $\forall x\in S$,因此,$\bar{x}$ 是 f 的全局极小点.

必要性 假定 $\bar{x}\in S$ 为 f 的全局极小点,构造 $A_1,A_2\in\mathbf{R}^{n+1}$,

其中:$A_1=\{((x-\bar{x}),\alpha)\mid x\in\mathbf{R}^n,\alpha>f(x)-f(\bar{x})\}$,

$A_2=\{((x-\bar{x}),\alpha)\mid x\in S,\alpha\leqslant 0\}$.

容易验证 A_1 和 A_2 都是凸集. 而且 $A_1\cap A_2=\varnothing$,因为否则存在一点 (x,α),使得

$$x\in S,\quad 0\geqslant\alpha>f(x)-f(\bar{x}).$$

这与 $\bar{x}$ 是 f 的全局极小点相矛盾,从而存在一个超平面分离 A_1 和 A_2;亦即,存在一个非零向量 (g_0,μ) 和一个数 β,使得

$$g_0^{\mathrm{T}}(x-\bar{x})+\mu\alpha\leqslant\beta,\quad x\in\mathbf{R}^n,\alpha>f(x)-f(\bar{x}),\quad (1-31)$$

$$g_0^{\mathrm{T}}(x-\bar{x})+\mu\alpha\geqslant\beta,\quad x\in S,\alpha\leqslant 0.\quad (1-32)$$

如果在(1-32)式中,令 $x=\bar{x}$ 和 $\alpha=0$,得到 $\beta\leqslant 0$. 在(1-31)式中,令 $x=\bar{x}$ 和 $\alpha=\varepsilon>0$,得到 $\mu\varepsilon\leqslant\beta$. 因为对每一个 $\beta>0$ 这都成立. 因面 $\mu\leqslant 0$ 且 $\beta\geqslant 0$.

概括起来,我们证明了 $\mu\leqslant 0$ 和 $\beta=0$. 下面我们接着证明 $\mu<0$.

假如 $\mu=0$,则由(1-31)式,对每个 $x\in\mathbf{R}^n$,$g_0^{\mathrm{T}}(x-\bar{x})\leqslant 0$,若令 $x=\bar{x}+g_0$,则有

$$0\geqslant g_0^{\mathrm{T}}(x-\bar{x})=\|g_0\|^2,$$

因此 $g_0=0$,而 $(g_0,\mu)\neq(0,0)$,故 $\mu<0$.

用 $-\mu$ 除(1-31)和(1-32),并用 g 表示 $-\dfrac{g_0}{\mu}$,就得出下面的不等式:

$$\alpha\geqslant g^{\mathrm{T}}(x-\bar{x}),\qquad x\in\mathbf{R}^n,\alpha>f(x)-f(\bar{x}),\quad (1-33)$$

$$g^{\mathrm{T}}(x-\bar{x})-\alpha\geqslant 0,\quad x\in S,\alpha\leqslant 0.\quad (1-34)$$

在(1-33)中,令 $\alpha=0$,得到对一切的 $x\in S$,$g^{\mathrm{T}}(x-\bar{x})\geqslant 0$.

再由(1-34),显然有 $f(x)\geqslant f(\bar{x})+g^{\mathrm{T}}(x-\bar{x})$,对一切 $x\in\mathbf{R}^n$,所以 g 是 $f(x)$ 在 $\bar{x}$ 处的次梯度,且对一切的 $x\in S$,$g^{\mathrm{T}}(x-\bar{x})\geqslant 0$.

对于定理 1.18，我们有如下两个有用的推论.

推论 1.1　在定理 1.18 的假设下，若 S 是开的，则 $\bar{x}$ 是 f 的全局极小点的必要充分条件是 $f(x)$ 在 $\bar{x}$ 处存在一个零次梯度.

证　由定理 1.18，$\bar{x}$ 是 f 的全局极小点的必要充分条件是对每一个 $x \in S$，有 $g^T(x-\bar{x}) \geq 0$，其中 g 是 $f(x)$ 在 $\bar{x}$ 处的次梯度. 因为 S 是开的，则存在一个 $\lambda > 0$，使得 $\bar{x} - \lambda g \in S$，因而 $-\lambda \|g\|^2 \geq 0$，从而 $\|g\|^2 \leq 0$，即 $g = 0$.

推论 1.2　在定理 1.18 的假设下，又假设 $f(x)$ 在 $\bar{x}$ 处可微，则 $\bar{x}$ 是 f 的全局极小点的必要充分条件是对一切的 $x \in S$，有 $\nabla f(\bar{x})^T(x-\bar{x}) \geq 0$. 此外，如果 S 是开的，$\bar{x}$ 是全局极小点的必要充分条件是 $\nabla f(\bar{x}) = 0$.

证　由定理 1.17、定理 1.18 及定理 1.18 的推论 1 易证.

定理 1.18 及其推论给出了刻画全局极小点的充要条件，具有重要的意义. 下面我们再给出两个刻画全局极小点的充分条件.

定理 1.19　设 $S \subset \mathbf{R}^n$ 凸，$f: S \to \mathbf{R}^1$ 在 int S 为连续可微凸函数，若 $x^* \in$ int S 为 $f(x)$ 的驻点，即 $\nabla f(x^*) = 0$，则 x^* 为 $f(x)$ 在 int S 上的全局极小点.

证　对 $\forall x \in$ int S 且 $x \neq x^*$，由定理 1.13，有

$$f(x) \geq f(x^*) + \nabla f(x^*)^T(x - x^*) = f(x^*),$$

从而 x^* 为 $f(x)$ 在 int S 上的全局极小点得证.

此定理说明了连续可微凸函数的驻点与全局极小点的关系，下面我们利用定理 1.14 和定理 1.19，导出定理 1.20.

定理 1.20　设 $S \subset \mathbf{R}^n$ 为凸集，$f: S \to \mathbf{R}^1$ 在 int S 二次连续可微，若 $x^* \in$ int S 为 $f(x)$ 的驻点，且对 $\forall x \in$ int S，$\nabla^2 f(x) \geq 0$，则 x^* 为 $f(x)$ 在 int S 上的全局极小点.

证　因 $f: S \to \mathbf{R}^1$ 在 int S 二次连续可微，且对 $\forall x \in$ int S，$\nabla^2 f(x) \geq 0$，由定理 1.14，知 $f(x)$ 在 int S 上的凸函数，再由定理 1.19，可知 x^* 为 $f(x)$ 在 int S 上的全局极小点.

1.4　拟凸函数与全局最优

定义 1.15　设 $S \subset \mathbf{R}^n$，非空凸，$f: S \to \mathbf{R}^1$，如果对 $\forall x_1, x_2 \in S$，$\lambda \in (0,1)$，有 $f(\lambda x_1 + (1-\lambda)x_2) \leq \max\{f(x_1), f(x_2)\}$，则称 $f(x)$ 是 S 上的拟凸函数；若对 $\forall x_1, x_2 \in S$，且 $f(x_1) \neq f(x_2)$，有 $f(\lambda x_1 + (1-\lambda)x_2) < \max\{f(x_1), f(x_2)\}$，则称 $f(x)$ 是 S 上的严格拟凸函数；如果对 $\forall x_1, x_2 \in S$，$\lambda \in (0,1)$，且 $x_1 \neq x_2$，有 $f(\lambda x_1 + (1-\lambda)x_2) < \max\{f(x_1), f(x_2)\}$，则

称 $f(\boldsymbol{x})$ 是 S 上的强拟凸函数.

显然,非空凸集 S 上的凸函数是拟凸函数,反之,则不成立.

拟凸函数与非拟凸函数在一维情形下,如图 1－12 所示:

图 1－12

前面,我们学习了凸函数可以用它的上图 epi f 的凸性来刻画.类似地,拟凸函数可以用它的水平集 S_α 来刻画,这一结果我们在定理 1.21 中给出.

定理 1.21 设 $S\subset\mathbf{R}^n$ 为非空凸集,$f:S\to\mathbf{R}^1$,函数 f 是拟凸的必要充分条件是 $S_\alpha=\{\boldsymbol{x}\in S;f(\boldsymbol{x})\leqslant\alpha\}$ 对每一个实数 α 是凸的.

证 必要性 假定 f 是拟凸的,并设 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S_\alpha$,因而 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S$ 且 $\max\{f(\boldsymbol{x}_1),f(\boldsymbol{x}_2)\}\leqslant\alpha$,对 $\forall\lambda\in(0,1)$ 以及 $\boldsymbol{x}=\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2$,由 S 的凸性,有 $\boldsymbol{x}\in S$.而且由 f 的拟凸性,知 $f(\boldsymbol{x})\leqslant\max\{f(\boldsymbol{x}_1),f(\boldsymbol{x}_2)\}$,从而 $f(\boldsymbol{x})\leqslant\alpha$,故 $\boldsymbol{x}\in S_\alpha$,于是 S_α 是凸的.

充分性 假定 S_α 对每个 α 是凸的,设 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S$,并设 $\lambda\in(0,1)$ 和 $\boldsymbol{x}=\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2$,若取 $\alpha=\max\{f(\boldsymbol{x}_1),f(\boldsymbol{x}_2)\}$,当 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S_\alpha$,由假定 S_α 是凸的,得知 $\boldsymbol{x}\in S_\alpha$,所以 $f(\boldsymbol{x})\leqslant\alpha=\max\{f(\boldsymbol{x}_1),f(\boldsymbol{x}_2)\}$,因此 f 是拟凸的.

下面我们给出一个刻画可微拟凸函数特征的必要充分条件.

定理 1.22 设 $S\subset\mathbf{R}^n$ 为非空凸集,$f:S\to\mathbf{R}^1$ 在 S 上可微,则 f 是拟凸的当且仅当下列等价陈述中任何一个陈述成立.

(1) 若 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S$ 且 $f(\boldsymbol{x}_1)\leqslant f(\boldsymbol{x}_2)$,则 $\nabla^{\mathrm{T}}f(\boldsymbol{x}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)\leqslant 0$;

(2) 若 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S$ 且 $\nabla^{\mathrm{T}}f(\boldsymbol{x}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)>0$,则 $f(\boldsymbol{x}_1)>f(\boldsymbol{x}_2)$.

证 显然陈述(1)和(2)是等价的,故只需证明 f 拟凸当且仅当陈述(1)成立.设 f 是拟凸的,并设 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S$ 使得 $f(\boldsymbol{x}_1)\leqslant f(\boldsymbol{x}_2)$,由 f 在 $\boldsymbol{x}_2$ 可微知,对 $\forall\lambda\in(0,1)$,有

$$f(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2)-f(\boldsymbol{x}_2)=\lambda\nabla f(\boldsymbol{x}_2)^{\mathrm{T}}(\boldsymbol{x}_1-\boldsymbol{x}_2)+\lambda\|\boldsymbol{x}_1-\boldsymbol{x}_2\|\alpha(\boldsymbol{x}_2,\lambda(\boldsymbol{x}_1-\boldsymbol{x}_2)),$$

其中 $\alpha(\boldsymbol{x}_2,\lambda(\boldsymbol{x}_1-\boldsymbol{x}_2))\to 0$,当 $\lambda\to 0$ 时,由 f 的拟凸性,我们有

$$f(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2)\leqslant f(\boldsymbol{x}_2),$$

因此,上述不等式蕴涵

$$\lambda \nabla f(\boldsymbol{x}_2)^{\mathrm{T}}(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \lambda \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \alpha(\boldsymbol{x}_2, \lambda(\boldsymbol{x}_1 - \boldsymbol{x}_2)) \leqslant 0,$$

除以 λ 并令 $\lambda \to 0$，就得到 $\nabla^{\mathrm{T}} f(\boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2) \leqslant 0$.

反之，假设 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ 使得 $f(\boldsymbol{x}_1) \leqslant f(\boldsymbol{x}_2)$，我们只需证明对 $\forall \lambda \in (0,1)$，$f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \leqslant f(\boldsymbol{x}_2)$，为此，证明集 L 是空的，

$$L = \{\boldsymbol{x} \mid \boldsymbol{x} = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2, \lambda \in (0,1), f(\boldsymbol{x}) > f(\boldsymbol{x}_2)\}.$$

反证，假设存在一个 $\boldsymbol{x}' \in L$，则 $\exists$ 一个 $\lambda \in (0,1)$ 且 $f(\boldsymbol{x}') > f(\boldsymbol{x}_2)$，使得

$$\boldsymbol{x}' = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2,$$

由于 f 是可微的，因而是连续的，于是存在一个 $\delta \in (0,1)$，使得

$$f(\mu \boldsymbol{x}' + (1-\mu)\boldsymbol{x}_2) > f(\boldsymbol{x}_2), \tag{1-35}$$

对 $\forall \mu \in [\delta, 1]$ 成立，以及 $f(\boldsymbol{x}') > f(\delta \boldsymbol{x}' + (1-\delta)\boldsymbol{x}_2)$，由此不等式及中值定理，我们应该有

$$0 < f(\boldsymbol{x}') - f(\delta \boldsymbol{x}' + (1-\delta)\boldsymbol{x}_2) = (1-\delta) \nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}' - \boldsymbol{x}_2), \tag{1-36}$$

其中，$\hat{\boldsymbol{x}} = \hat{\mu}\boldsymbol{x}' + (1-\hat{\mu})\boldsymbol{x}_2$，对某个 $\hat{\mu} \in (\delta, 1)$，由 (4－1)，显然 $f(\hat{\boldsymbol{x}}) > f(\boldsymbol{x}_2)$，

用 $1-\delta > 0$ 除(1－36)得到 $\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}' - \boldsymbol{x}_2) > 0$，这就蕴涵了

$$\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}_1 - \boldsymbol{x}_2) > 0. \tag{1-37}$$

但是，另一方面，$f(\hat{\boldsymbol{x}}) > f(\boldsymbol{x}_2) \geqslant f(\boldsymbol{x}_1)$ 和 $\hat{\boldsymbol{x}}$ 是 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 的一个凸组合，由定理的假设，$\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}_1 - \hat{\boldsymbol{x}}) \leqslant 0$，于是我们就有：(假定 $\hat{\boldsymbol{x}} = \hat{\lambda}\boldsymbol{x}_1 + (1-\hat{\lambda})\boldsymbol{x}_2$，其中 $\hat{\lambda} \in (0,1)$)

$$0 \geqslant \nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}_1 - \hat{\boldsymbol{x}}) = (1-\hat{\lambda}) \nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}_1 - \boldsymbol{x}_2),$$

此不等式与(1－37)是不相容的，故 L 是空的.

为了说明上述定理，设 $f(x) = x^3$，为了检查 f 的拟凸性，假定 $f(x_1) \leqslant f(x_2)$，即 $x_1^3 \leqslant x_2^3$. 只有当 $x_1 \leqslant x_2$ 时才成立. 现在考虑 $\nabla f(x_2)(x_1 - x_2) = 3(x_1 - x_2)x_2^2$，因为 $x_1 \leqslant x_2$，所以 $3(x_1 - x_2)x_2^2 \leqslant 0$，因此 $f(x_1) \leqslant f(x_2)$ 蕴涵着 $\nabla f(x_2)(x_1 - x_2) \leqslant 0$，由定理知 f 是拟凸的. 下面我们看另一个例子，设 $f(x_1, x_2) = x_1^3 + x_2^3$，令 $\boldsymbol{y}_1 = (2, -2)^{\mathrm{T}}$ 和 $\boldsymbol{y}_2 = (1, 0)^{\mathrm{T}}$，注意到 $f(\boldsymbol{y}_1) = 0$ 和 $f(\boldsymbol{y}_2) = 1$，从而 $f(\boldsymbol{y}_1) \leqslant f(\boldsymbol{y}_2)$，但是，另一方面 $\nabla f(\boldsymbol{y}_2)^{\mathrm{T}}(\boldsymbol{y}_1 - \boldsymbol{y}_2) = (3, 0)(1, -2)^{\mathrm{T}} = 3$. 由定理的必要部分，$f$ 不是拟凸的. 这也证明了两个拟凸函数之和不一定是拟凸的.

下面我们来证明严格拟凸函数能保证凸集上的局部极小点为全局极小点.

定理 1.23　设 $f: \mathbf{R}^n \to \mathbf{R}^1$ 为严格拟凸函数，$S \subset \mathbf{R}^n$ 为一非空凸集，如

果 $\bar{x}$ 是 f 在 S 上的局部极小点，则 $\bar{x}$ 也是 f 在 S 上的全局极小点.

证　反证，若 $\bar{x}$ 不是 f 在 S 上的全局极小点，则 $\exists$ 一个 $\hat{x} \in S$，有 $f(\hat{x}) < f(\bar{x})$，由 S 的凸性，对 $\forall \lambda \in (0,1)$，$\lambda\hat{x} + (1-\lambda)\bar{x} \in S$. 根据假设，因 $\bar{x}$ 是局部极小点，故不存在一个 $\delta \in (0,1)$，对一切 $\lambda \in (0,\delta)$，有 $f(\bar{x}) \leqslant f(\lambda\hat{x} + (1-\lambda)\bar{x})$. 但是，由于 f 是严格拟凸的，并且 $f(\hat{x}) < f(\bar{x})$，则对 $\forall \lambda \in (0,1)$，有 $f(\lambda\hat{x} + (1-\lambda)\bar{x}) < f(\bar{x})$，从而产生矛盾. 这就说明 $\hat{x}$ 不存在. 故 $\bar{x}$ 是 f 在 S 上的全局极小点.

最后，我们来证明强拟凸函数能保证凸集上的局部极小点为唯一的全局极小点.

定理 1.24　设 $f:\mathbf{R}^n \to \mathbf{R}^1$ 为强拟凸函数，$S \subset \mathbf{R}^n$ 为一非空凸集，如果 $\bar{x}$ 是 f 在 S 上的局部极小点，则 $\bar{x}$ 是 f 在 S 上的唯一全局极小点.

证　因 $\bar{x}$ 是一个局部极小点，则在 x 附近存在一个 ε - 邻域 $N_\varepsilon(\bar{x})$，使得对一切

$$x \in S \cap N_\varepsilon(\bar{x}), \quad f(\bar{x}) \leqslant f(x).$$

反证法，假设定理的结论不成立，则存在一点 $\hat{x} \in S$，使得 $\hat{x} \neq \bar{x}$ 且 $f(\hat{x}) \leqslant f(\bar{x})$.

由强拟凸性，对一切 $\lambda \in (0,1)$，得出

$$f(\lambda\hat{x} + (1-\lambda)x) < \max\{f(\hat{x}), f(\bar{x})\} = f(\bar{x}).$$

但当 λ 足够小，$(\lambda\hat{x} + (1-\lambda)\bar{x}) \in S \cap N_\varepsilon(\bar{x})$. 于是上面的不等式与 $\bar{x}$ 的局部最优性相矛盾. 证明完成.

注　每个强拟凸函数都是严格拟凸的，故在定理 1.23、定理 1.24 中的 $\bar{x}$ 显然是 f 在 S 上的全局极小点.

第 2 章　最优化方法概述

2.1　最优化问题的提法及分类

最优化问题通常可以表为函数的极值问题，即

$$\min_{x \in \mathbf{R}^n} f(x), \tag{2-1}$$

其中，$f: S \subset \mathbf{R}^n \to \mathbf{R}^1$ 为实值函数，称为目标函数，S 为 $f(x)$ 的定义域．也可以写成更一般的形式：

$$\min_{x \in S} f(x), \tag{2-2}$$

$$\text{s.t.} \quad \varphi(x) = 0,$$
$$\psi(x) \leqslant 0,$$

其中

$$x = (x_1, \cdots, x_n)^{\mathrm{T}},$$
$$\varphi(x) = (\varphi_1(x), \cdots, \varphi_n(x))^{\mathrm{T}},$$
$$\psi(x) = (\psi_1(x), \cdots, \psi_m(x))^{\mathrm{T}}.$$

在问题(2－1)中，自变量 x 可以通取定义域 S 内的值而不受限制，故称为无约束极值问题，而在问题(2－2)中 $x \in S$ 必须受到一定的约束，即

$$\varphi_i(x) = 0, \quad i = 1, \cdots, n, \tag{2-3}$$
$$\psi_j(x) \leqslant 0, \quad j = 1, \cdots, m, \tag{2-4}$$

故称(2－2)为约束极值问题．(2－3)、(2－4)都称为约束条件，其中(2－3)是等式约束，(2－4)是不等式约束．也称 $x \in S$ 为集约束，在我们的问题中，集约束是无关紧要的，这是因为有时 $S \equiv \mathbf{R}^m$，不然的话，S 通常也可以用不等式约束表达出来．因此，今后一般不再考虑集约束．

上述问题又统称为数学规划．按照目标函数以及约束条件中约束函数的类型和自变量取值状态．还可以区分为不同形式的数学规划．

当 $f(x)$，$\varphi_i(x)$，$\psi_j(x)$ 均为线性函数时，称为线性规划，当其中任一个为非线性函数时，称为非线性规划；当 $f(x)$ 为二次函数，$\varphi_i(x)$，$\psi_j(x)$ 均为线性函数时，称为二次规划；当 x 的各个分量全部或部分只取离散值(如取整数值)称为整数规划；在整数规划中，若 x 的各个分量只取 0，1 两个值时，称为 0－1 规划．还有几何规划，动态规划等等．

求解问题(2-1)或(2-2)的目的，是要确定一个向量 x^*，满足

$$f(x^*) = \min_{x\in \mathbf{R}^n} f(x),$$

对于问题(2-2)，还需满足

$$\varphi(x^*) = 0, \qquad \psi(x^*) \leqslant 0,$$

我们称 x^* 为最优点．对于极大问题

$$\max F(x),$$

如果令 $f(x) = -F(x)$，则转化为极小问题

$$\min f(x),$$

因此，往后我们都是研究极小问题，但不失其普遍意义．

2.2　最优化问题举例

例 2.1　产品生产过程设计

设某厂用一定的生产流程成批生产某种化工产品，在保证质量的前提下我们把单位重量的产品价格作为评价的指标．显然，影响指标因素首先是原料的总量，同时还有生产过程中的有关因素：温度、压力等等．于是根据经验或理论分析，我们可以把评价指标(价格)表为若干个参数 x_1，…，x_n 的函数

$$f(x_1,\cdots,x_n),$$

并且我们所关心的问题是：求 $x_1,\cdots,x_n$，使 $f(x_1,\cdots,x_n)$ 达到极小．

从工程上或经济上来考虑，$x_1,\cdots,x_n$ 的取值不能随意，必须按照实际给予约束，最简单的约束是

$$a_i \leqslant x_i \leqslant b_i \qquad (i = 1,\cdots,n),$$

$$S = \{x \mid a_i \leqslant x_i \leqslant b_i, i = 1,\cdots,n\},$$

更一般的约束是

$$\varphi_i(x_1,\cdots,x_n) = 0 \quad (i = 1,\cdots,n),$$

$$\psi_i(x_1,\cdots,x_m) \leqslant 0 \quad (j = 1,\cdots,m).$$

如果不对 x 以限制，则形成问题(2-1)，若对 x 加上上述限制，则形成问题(2-2)．

在生产实践和科学实践中，与例 2.1 类似的实际问题大量存在．一类问题是工业设计和控制中的非线性分析，例如：结构系统最优设计(两杆桁架——人字架设计)；机械零件或部件的最优设计(齿轮轴颈、凸轮设计)；化工设备最优设计(单件或连锁设备的优化设计)；电力网络和水力网络的优化设计(平衡条件)等等．上述问题通常表示为较为复杂的非线

性规划问题.另一类问题是生产计划和经营决策中的优化问题.例如:合理计划生产问题,包括运输、分配、布局、选址、指派下料、配料等优化问题.这些问题通常用线性规划来表示;合理开发(或配置)资源问题,包括可再生资源的持续开发,不可再生资源的优化配置等.这些问题通常用非线性规划来表示;合理组合投资问题,即追求最大收益、最小风险的投资组合方案.这类问题通常用多目标规划来表示.

例 2.2 模型参数识别

在某些学科的研究中,根据学科的有关知识,假定已经知道在一定范围内控制所研究现象的规律,即该现象的数学模型已被给定.例如是时间的函数,即

$$f(t) = \theta_1 + \theta_2 e^{-\theta_3 t}, \tag{2-5}$$

实验室进行试验的目的,是要根据在时刻 $t_1, t_2, \cdots, t_m$ 所测得的 f 值即 $f_1, \cdots, f_m$ 来估计(2-5)中未知参数 $\theta_1, \theta_2, \theta_3$ 的值.

这里的决策就是合理地指定这些参数的值,而合理的标准即是使这些值在某种意义下为最优.因此问题的关键是选定目标函数.例如我们采用最小二乘法作偏差平方和函数:

$$s(\theta_1, \theta_2, \theta_3) = \sum_{i=1}^{m} (f_i - f(t_i))^2 = \sum_{i=1}^{m} (f_i - (\theta_1 + \theta_2 e^{-\theta_3 t_i}))^2, \tag{2-6}$$

于是问题变成确定 $\theta_1, \theta_2, \theta_3$ 使 $s(\theta_1, \theta_2, \theta_3)$ 达到极小.由于对模型(2-6)中参数 $\theta_1, \theta_2, \theta_3$ 没有任何附加条件,故形成无约束最优化问题.

如果考虑到这些参数的物理意义,那么作为无约束问题解得的参数,可能不能接受.为了防止这种情况,宁愿在数学模型中把对参数值的限制也包括进去.例如要求 $\theta_3 \geq 0$ 以及 $f(0) = 1$,即 $\theta_1 + \theta_2 = 1$,于是形成约束问题:

$$\begin{cases} \min\ s(\theta_1, \theta_2, \theta_3), \\ \text{s.t.}\quad \theta_1 + \theta_2 = 1, \\ \qquad\quad \theta_3 \geq 0. \end{cases} \tag{2-7}$$

例 2.3 曲线拟合问题

这种问题与模型的参数识别有紧密的关系,只是问题的提法不同而已.

设给出观测数据(即观察函数 $y = y(t)$ 的实测值):

$$(t_i, y_i), \qquad i = 1, \cdots, m,$$

需要确定 y 与 t 之间的经验公式.

与例 2.2 不同的是对于 y 与 t 之间的关系没有经验可借鉴. 因此，数学模型的选择是自由的. 通常选取一个独立的函数系 $\{\varphi_i(t)\}_{i=1,\cdots,n}(n<m)$，对 $t\in[a,b]$，函数系

$$\varphi_1(t),\varphi_2(t),\cdots,\varphi_n(t)$$

是线性独立的. 例如，常用的函数系有：

幂函数系 $\{t^{j-1}\}$：$1,t^1,t^2,\cdots,t^{n-1}$；

三角函数系 $\{\sin jt\}$：$\sin t,\sin 2t,\cdots,\sin nt$；

指数函数系 $\{e^{\lambda_j t}\}$：$e^{\lambda_1 t},e^{\lambda_2 t},\cdots,e^{\lambda_n t}$.

令

$$y(t)=\sum_{i=1}^{n}x_i\varphi_i(t), \tag{2-8}$$

其中：$x_1,x_2,\cdots,x_n$ 为模型的待定参数，仍采用例 2.2 的做法，作偏差平方和函数

$$s(x_1,\cdots,x_n)=\sum_{i=1}^{m}(y(t_i)-y_i)^2=\sum_{i=1}^{m}\Big(\sum_{j=1}^{n}x_j\varphi_j(t_i)-y_i\Big)^2, \tag{2-9}$$

并确定 $x_1^*,\cdots,x_n^*$，使

$$s(x_1^*,\cdots,x_n^*)=\min s(x_1,\cdots,x_n).$$

例 2.2 和例 2.3 中所示的问题在各种工程技术中都有大量的应用. 除上述三类实际问题外，许多数学问题的求解也常常归结为最优化问题. 例如非线性方程组的求解，常微分方程边值问题的求解，变分问题的求解等等.

虽然实际问题中多数极值问题都是以约束问题的形式出现，但是我们研究无约束极值问题的解法仍然是有现实意义的. 这是因为：

(1) 如例 2.2 和例 2.3 所示，实际中存在大量的平方和极小化问题，求这类问题不带约束的更为普遍；

(2) 某些求解约束极值的算法是以无约束极值算法为基础的.

正因为如此，近二十年来，人们极为重视无约束问题算法的研究，做出了很多有意义的成果. 读者可参阅附录Ⅱ——最优化方法的发展进程.

以上叙述的多元函数的极值问题又称为静态优化问题. 其特点是问题中的目标函数 $f(x)$ 是一个实值函数，决策变量 x 不含时变. 最优化问题的另一种类型是研究泛函极值，它是最优控制讨论的对象，其中目标函数是一个泛函，状态变量 $x(t)$，控制变量 $u(t)$ 都是定义在一个时间区间上的函数. 由于这类问题含时变，通常又称为动态优化问题.

例 2.4（连续最优控制问题）

考虑一电梯从某一高度最速下降到地面(如图 2－1).

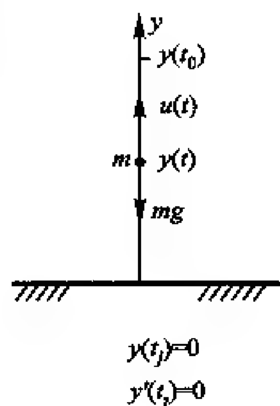


图 2－1

设 $y(t)$ 表示时刻 t 电梯的高度(又称为状态函数),$u(t)$ 表示时刻 t 电梯承受的推力(控制函数),m 表示电梯的质量.

根据牛顿第二定律,容易得到这个系统的状态方程

$$my''(t) = u(t) - mg,$$

即

$$y''(t) + g = \frac{1}{m}u(t), \qquad (2-10)$$

问题目标归结为电梯从初态 $y(t_0)$ 出发,最速下降至地面,即终态 $y(t_f)=0$. 也就是说,要求运动的控制时间为最短:

$$J = t_f - t_0 = \int_{t_0}^{t_f} \mathrm{d}t = \min, \qquad (2-11)$$

因此,上述问题的数学模型为

$$\min J = \int_{t_0}^{t_f} \mathrm{d}t$$

$$\text{s.t.} \begin{cases} y''(t) + g = \dfrac{1}{m}u(t); \\ |u(t)| < k. \end{cases} \qquad (2-12)$$

其中常数 k 是推力的一个上界.

若令 $y_1(t)=y(t)$,$y_2(t)=y'(t)$,则(2－12)分解为一阶方程组

$$\begin{cases} y_1'(t) = y_2(t), \\ y_2'(t) = \dfrac{1}{m}u(t) - g. \end{cases}$$

初始状态为

$$y_1(t_0) = y(t_0), y_2(t_0) = y'(t_0).$$

终端状态为

$$y_1(t_f) = y(t_f) = 0, y_2(t_f) = y'(t_f) = 0.$$

与上述最速升降问题类似的实际问题大量存在. 如生产分配与生产计划问题;反应器控制与液体搅拌问题;空间技术中的最小耗能问题、登月问题、防天拦截问题等等,这一类问题的共同提法是:在一定条件下(状态约束、控制约束),选择可控制的参数,使某一连续过程从初始时间和初始状态出发,在最终时刻最优地达到预计的目标. 其数学模型为

$$\min J(\boldsymbol{u}) = \int_{t_0}^{t_f} f_0(t, \boldsymbol{x}(t), \boldsymbol{u}(t))\,dt.$$

$$\text{s. t.} \begin{cases} \boldsymbol{x}'(t) = \boldsymbol{f}(t, \boldsymbol{x}(t), \boldsymbol{u}(t)), t \in [t_0, t_f], \\ \boldsymbol{x}(t) \in \boldsymbol{X} \subset \mathbf{R}^n, \\ \boldsymbol{u}(t) \in \boldsymbol{U} \subset \mathbf{R}^p, \end{cases} \qquad (2-13)$$

其中初始状态 $\boldsymbol{x}(t_0)$ 为已知，终端状态 $\boldsymbol{x}(t_f)$ 自由或者受约束.

满足约束条件的 $\boldsymbol{x}(t)$，$\boldsymbol{u}(t)$ 称为问题(2－13)的容许解($\boldsymbol{x}(t)$ 称为容许状态，$\boldsymbol{u}(t)$ 称为容许控制).

使 $\min J(\boldsymbol{u}(t)) = J(\boldsymbol{u}^*(t))$ 的容许解 $\boldsymbol{u}^*(t)$，$\boldsymbol{x}^*(t)$ 称为问题(2－13)的最优解($\boldsymbol{u}^*(t)$ 为最优控制，$\boldsymbol{x}^*(t)$ 为相应的最优轨迹).

动态优化问题在一定条件下可以转化为静态优化问题来求解. 例如连续最优控制问题经离散化后，可以化为一个非线性规划问题. 因此，本书可为读者进一步学习最优控制的计算方法奠定基础.

2.3 无约束极值问题算法综述

本节概括地讨论无约束极值问题算法——搜索算法的特征. 设无约束极小值问题

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} f(\boldsymbol{x}), \qquad (2-14)$$

对于正定二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c, \qquad (2-15)$$

由于 $\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}$ 正定，据第 1 章定理 1.14，$f(\boldsymbol{x})$ 在 $\mathbf{R}^n$ 上存在全局极小点 $\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$，即 $f(\boldsymbol{x})$ 的稳定点，它满足 $\nabla f(\boldsymbol{x}^*) = 0$.

对于一般的非线性函数 $f(\boldsymbol{x})$，寻求极小点就困难得多，古典分析的方法是：首先利用必要条件 $\nabla f(\boldsymbol{x}) = 0$ 求出稳定点 $\boldsymbol{x}^*$，然后利用充分条件(第 1 章定理 1.11)及凸性条件(第 1 章定理 1.14—定理 1.16)判别 $\boldsymbol{x}^*$ 是否为问题(2－14)的极小点.

然而，用古典分析的方法求解问题(2－14)实际上是很难行通的，这是由于：

(1) 实际中相当数量的目标函数 $f(\boldsymbol{x})$ 不具有解析性，因此非线性方程组 $\nabla f(\boldsymbol{x}) = 0$ 无法形成；

(2) 即使形成了 $\nabla f(\boldsymbol{x}) = 0$，求它的解几乎与解决原问题一样地困难；

(3) 即使求得了 $\nabla f(\boldsymbol{x}) = 0$ 的解 $\boldsymbol{x}^*$，但由于 $f(\boldsymbol{x})$ 的上述充分条件或

凸性条件不被满足或难于验证，因此仍无法确定 x^* 是否为问题(2－14)的解.

鉴于上述种种原因，对于问题(2－14)的求解，通常采用搜索算法，即下降算法. 并且根据目标函数解析性的好坏，把下降算法分为两类：一类是解析方法，即在构造算法时，用到了 $f(x)$ 的导数值(或者一阶导数 $\nabla f(x)$，或者二阶导数 $H(x)=\nabla^2 f(x)$)；另一类是直接法，构造算法仅用到了 $f(x)$ 的值.

2.3.1　下降算法

在定义了下降方向(第 1 章定义 1.11)后便可构造下降算法. 对于给定的初始点 x_0，作如下计算：

(1) 取定下降方向 p_0：使 p_0 满足 $\nabla f(x)^T p_0<0$；

(2) 确定步长因子 α_0：即以 p_0 为方向，作射线 $x=x_0+\alpha p_0$，在此射线上(图 2－2)定出 $\alpha=\alpha_0$ 使得：

$$f(x_0+\alpha_0 p_0)<f(x_0);$$

(3) 迭代：以 $x_1=x_0+\alpha p_0$ 作为 x^* 的第一级近似值.

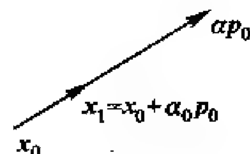


图 2－2

重复上述步骤，得到一般的迭代公式：

$$x_{k+1}=x_k+\alpha_k p_k,$$

其中：下降方向 p_k 满足　$\nabla f(x_k)^T p_k<0$；

步长因子 α_k 满足　$f(x_k+\alpha p_k)<f(x_k)$，$k=0,1,\cdots$.

综上所述，给出下降算法的一般步骤.

算法 2.1

① 给定极小点 x^* 的初始近似值 x_0；

② 置 $k=0$；

③ 计算 p_k：$\nabla f(x_k)^T p_k<0$；

④ 计算 α_k：$f(x_k+\alpha_k p_k)<f(x_k)$；

⑤ 计算 $x_{k+1}=x_k+\alpha_k p_k$；

⑥ 检查终止法则，若 x_{k+1} 达到收敛精度，则输出 x_{k+1}，停止迭代；

⑦ 置 $k=k+1$ 转③.

以上定义的下降算法，它具有迭代法的特征，下降方向(或称搜索方向，寻查方向)p_k 和步长因子 α_k 构成每一次迭代的修正量，因此，它们是决定算法好坏的重要因素.

关于搜索方向 p_k 的选择，应遵循如下原则：

(1) 它必须是下降方向；

(2) 希望它尽可能指向极小点，或者使函数值下降最快；

(3) 构造它时，计算量不是太大.

例 2.5 设$f(x)$为正定二次函数，即

$$f(x)=\frac{1}{2}x^{T}Ax-b^{T}x+c,$$

其中 A 为正定对称矩阵. 在二维情形，$f(x)$的等高线 $f(x)=\gamma$ 是一族同心椭圆，其共同的中心为二次函数的极小点，如图 2－3 所示.

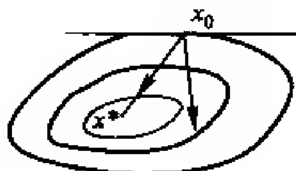


图 2－3

从任一初始点 x_0 出发，若选取负梯度方向，即最速下降方向 $p_0=-\nabla f(x_0)$进行迭代，不可能一次迭代达到$f(x)$的极小点，即使经过若干次迭代，也只能得到极小点的一个近似值，从例 2.6 中我们将会验证这一结论. 若将负梯度方向作线性变换，从任一初始点 x_0 出发，取下降方向 $p_0=-A^{-1}\nabla f(x_0)$进行迭代，则一次迭代便可达到极小点. 我们称 $p_0=-A^{-1}\nabla f(x_0)$为$f(x)$在 x_0 处的 Newton 方向.

事实上，如前所述，正定二次函数$f(x)$在 $\mathbf{R}^n$ 上有全局极小点 $x^*=A^{-1}b$，现取 $p_0=-A^{-1}\nabla f(x_0)$，并取 $\alpha_0=1$ 进行迭代，只要迭代一步便可达极小点，即

$$\begin{aligned}x_1&=x_0-A^{-1}\nabla f(x_0)=x_0-A^{-1}(Ax_0-b)\\&=x_0-x_0+A^{-1}b=A^{-1}b=x^*.\end{aligned}$$

由此可以看出，不同的下降方向之间仍有一个优劣比较的问题.

从局部看问题，在 x_0 点邻近，沿下降方向 $p=-\nabla f(x_0)$进行搜索，由于它是最速下降方向，函数值下降最快；但从全局看问题，沿 Newton 方向 $p_0=-A^{-1}\nabla f(x_0)$进行搜索，一次迭代便达极小点，用原则中的第二条衡量，它比 $-\nabla f(x_0)$好.

虽然这个例子告诉我们，对于正定二次函数，用 Newton 方向进行搜索是很理想的，但是构成 $p_0=-A^{-1}\nabla f(x_0)$却要花费太大的计算量，因此从原则的第三条来看，Newton 方向还是有缺陷的. 至于对一般的非线性函数$f(x)$，Newton 方向的缺陷更加突出. 我们将在第 4 章中详细讨论该方向的构造及性质.

综上所述，对于搜索方向的选择，尽管我们规定了三条原则，但不可能找到一个万全的算法，在选择方向时要全面进行考察(从函数的性态到三条原则)，而且各种算法也是互相补充的.

关于步长因子 α_k 的选择，有如下三种方案：

(1) 定步长，即取 $\alpha_k = 1$，由此形成的算法称为简单算法，但由第 1 章定理 1.9 可知，这种算法虽然计算方便，但不一定能保证算法的下降性质；

(2) 变步长，即每次迭代中，按如下原则确定 α_k：

$$f(\boldsymbol{x}_{k+1}) \leqslant f(\boldsymbol{x}_k),$$

其中 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$，由此形成的算法称为可接受点算法. 这种算法保证了每次迭代时函数值下降，但并非最优；

(3) 最优步长，即令 α_k 满足

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) = \min f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k),$$

即是说，$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$ 是 $f(\boldsymbol{x})$ 沿 $\boldsymbol{p}_k$ 方向上的极小点，由此形成的算法称为完备算法. 该算法的特点是，把一个多维极值问题简化为逐次沿 $\boldsymbol{p}_k$ 方向的一维搜索（一维极值）. 其优点是，它不仅能保证算法的下降性，而且还简化了算法的收敛条件. 往后我们知道，如果算法仅仅满足下降性质，还不能保证点列 $\{\boldsymbol{x}_k\}$ 收敛于问题(2-14)的极小点 $\boldsymbol{x}^*$，甚至不能保证收敛于 $f(\boldsymbol{x})$ 的稳定点. 因此，为了保证算法的收敛性质，还需对 $f(\boldsymbol{x})$ 提出一系列条件，而对于这些条件的简化将使得算法更有价值.

完备算法虽然优点较多，但也有一个严重的缺陷，即每进行一次一维搜索时需要较大的计算量，对于高维极值问题尤为突出. 20 世纪 60 年代以来，A. A. Goldstern (1962, 1965)，W. I. Zangwill (1969)，M. A. Wolfe (1969, 1971) 等人致力于研究不用一维搜索的算法（即可接受点算法）. 在对 $f(\boldsymbol{x})$ 要求不是太大的情况下保证由此所产生的点列 $\{\boldsymbol{x}_k\}$ 至少收敛于 $f(\boldsymbol{x})$ 的稳定点. 这些最新理论成果的详细讨论可参阅文献[1].

下面我们给出一个简单而又常用的下降算法，即梯度法，又称最速下降法.

算法 2.2

已知：目标函数 $f(\boldsymbol{x})$ 及其梯度 $\boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})$.

① 给定初始近似 $\boldsymbol{x}_0$；

② 计算 $f_0 = f(\boldsymbol{x}_0)$，$\boldsymbol{g}_0 = \boldsymbol{g}(\boldsymbol{x}_0)$；

③ 置 $k = 0$；

④ 计算负梯度方向 $\boldsymbol{p}_k = -\boldsymbol{g}_k$；

⑤ 计算 α_k：令 α_k 满足：

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) = \min f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k),$$

或者，令 α_k 满足：

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) \leqslant f(\boldsymbol{x}_k);$$

⑥ 计算 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$；

⑦ 计算 $f_{k+1}=f(\boldsymbol{x}_{k+1}),\boldsymbol{g}_{k+1}=\boldsymbol{g}(\boldsymbol{x}_{k+1})$;

⑧ 检查终止法则,若 $\boldsymbol{x}_{k+1}$ 达到收敛精度,则输出 $\boldsymbol{x}_{k+1}$,否则置 $k=k+1$ 转④.

在算法 2.2 中,根据确定 α_k 的不同方式,我们称算法 2.2 或者为梯度法的完备算法,或者为梯度法的可接受点算法,对于完备算法,若目标函数为正定二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c \quad (\boldsymbol{A} > 0), \tag{2-16}$$

可以导出 α_k 的显式表达式

$$\alpha_k = \frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k}. \tag{2-17}$$

公式(2-17)的推导过程将在第 3 章给出.

例 2.6 试用梯度法的完备算法求函数

$$f(x_1,x_2) = x_1^2 + 4x_2^2$$

的极小点. 设初始点为 $\boldsymbol{x}_0=(1,1)^{\mathrm{T}}$.

显然,极小点 $\boldsymbol{x}^*=(0,0)^{\mathrm{T}}$,现在我们用算法 2.2 来逼近这个极小点.

解 目标函数 $f(x_1,x_2)$ 的二次型矩阵为

$$\boldsymbol{A} = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix},$$

梯度表达式为

$$\boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 8x_2 \end{bmatrix},$$

由 $\boldsymbol{x}_0=(1,1)^{\mathrm{T}}$,得

$$\boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0) = (2,8)^{\mathrm{T}},$$

$$\boldsymbol{p}_0 = -\boldsymbol{g}_0 = -(2,8)^{\mathrm{T}}$$

$$\alpha_0 = \frac{\boldsymbol{g}_0^{\mathrm{T}}\boldsymbol{g}_0}{\boldsymbol{g}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_0} = 0.130\,77,$$

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0\boldsymbol{p}_0 = (0.738\,46, -0.046\,16)^{\mathrm{T}}.$$

又由 $\boldsymbol{x}_1$ 得

$$\boldsymbol{g}_1 = \nabla f(\boldsymbol{x}_1) = (1.476\,92, -0.369\,23)^{\mathrm{T}},$$

$$\boldsymbol{p}_1 = -\boldsymbol{g}_1 = (-1.476\,92, 0.369\,23)^{\mathrm{T}},$$

$$\alpha_1 = \frac{\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{g}_1}{\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_1} = 0.425\,00,$$

$$\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1\boldsymbol{p}_1 = (0.110\,76, 0.110\,76)^{\mathrm{T}}.$$

重复上述计算得

$$x_3 = (0.08162, -0.00579)^T.$$

由此看出，即使对于正定二次函数，由梯度法所产生的迭代点列 $\{x_k\}$ 收敛于 x^* 的速度也是慢的. 这个例子说明负梯度方向，即最速下降方向，它是在一个点邻近的局部性质，从全局来讲并非是最好的下降方向. 对于一个下降算法，如何规定其收敛速度及终止法则，我们将在 2.3.2 中详细介绍.

2.3.2 算法收敛速度及终止法则

搜索算法是一种迭代方法，如同任何一个迭代程序一样，我们必须考虑它的收敛条件，收敛速度以及终止法则. 对于收敛条件的讨论将在 2.3.3 中结合具体的下降算法进行. 本节将给出收敛速度的概念，并且采用事后估计的方法给出一般下降算法的终止法则.

1. 收敛速度

定义 2.1 在线性赋范空间 $\{\mathbf{R}^n, \|\cdot\|\}$ 中，设点列 $\{x_k\}$ 收敛于 x^*，且满足

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} = c, \tag{2-18}$$

其中 $p \geqslant 1$ 为实数，$c > 0$ 为与 k 无关的常数，则称 $\{x_k\}$ 为 p 阶收敛. 当 $p = 1$ 且 $0 < c < 1$ 时，$\{x_k\}$ 称为线性收敛.

当 $p > 1$，称 $\{x_k\}$ 为超线性收敛. 特别 $p = 2$ 时，$\{x_k\}$ 称为平方收敛（或称二阶收敛）.

在以后介绍的解析算法中，如梯度法具线性收敛速度，Newton 法具二阶收敛速度，共轭方向法（共轭梯度）则具超线性收敛速度.

根据收敛阶的定义，可以推导以下的定理，它对于构造迭代终止法则是有用处的.

定理 2.1 在线性赋范空间 $\{\mathbf{R}^n, \|\cdot\|\}$ 中，如果 $\{x_k\}$ 为超线性收敛 $(p > 1)$，则

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} = 1, \tag{2-19}$$

证 由收敛阶的定义，若 $p > 1$，则

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

另一方面

$$\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}=\frac{\|(x_{k+1}-x_k)+(x_k-x^*)\|}{\|x_k-x^*\|}$$

$$\geqslant\frac{\left|\|x_{k+1}-x_k\|-\|x_k-x^*\|\right|}{\|x_k-x^*\|}$$

$$=\left|\frac{\|x_{k+1}-x_k\|-\|x_k-x^*\|}{\|x_k-x^*\|}\right|$$

$$=\left|\frac{\|x_{k+1}-x_k\|}{\|x_k-x^*\|}-1\right|$$

所以

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x_k\|}{\|x_k-x_k^*\|}=1.$$

不难验证定理 2.1 的逆命题不必真.如果我们知道$\{x_k\}\to x^*$的收敛阶$p>1$,那么(2－19)成立,因此对于充分大的正整数k,可以用$\|x_{k+1}-x_k\|$来估计误差$\|x_k-x^*\|$,并且指望用$\|x_{k+1}-x_k\|<\varepsilon$(给定的精度)来终止迭代.

2. 迭代终止法则:下降算法所产生的点列$\{x_k\}$是一个无穷点列,当其点列收敛时,需要给出一个终止迭代的法则,即是说给出收敛的准则.

(1) 如前所述,当$\{x_k\}$的收敛阶$p>1$时,根据(2－19)式,可以用$\|x_{k+1}-x_k\|$代替$\|x_k-x^*\|$来估计误差.从而对预先给定的精度$\varepsilon<0$,以$\|x_{k+1}-x_k\|<\varepsilon$作迭代终止条件是适当的.但是单独使用上式是有毛病的.这是因为:第一,在实际计算时我们并不预先知道$\{x_k\}$的收敛阶;第二,$\|x_{k+1}-x_k\|$未必是k的单调减函数,在非超线性收敛的情形,会出现异常.图 2－4 表明,对于$f:\mathbf{R}^1\to\mathbf{R}^1$有可能出现这种情形,虽然$\|x_{k+1}-x_k\|$很小,但$|f_{k+1}-f_k|$仍较大,并且$x_k$远离极小点$x^*$.事实上,因为

$$\|x_{k+1}-x_k\|\leqslant\|x_{k+1}-x^*\|+\|x^*-x_k\|,$$

由$\|x_k-x^*\|$很小(k充分大),可以保证$\|x_{k+1}-x_k\|$很小,但是$\|x_{k+1}-x_k\|$很小并不能保证$\|x_k-x^*\|$很小.

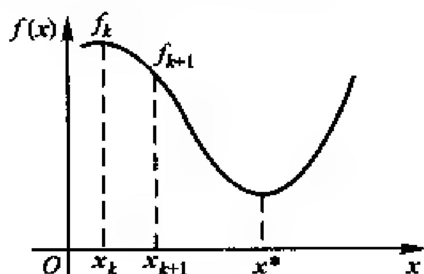


图 2－4

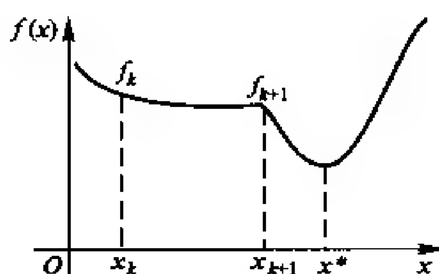


图 2－5

(2) 由于$\{f_k\}$是一个单调减序列,因此,当 k 增大时,$|f_{k+1}-f_k|$减小,从而可用$|f_{k+1}-f_k|<\varepsilon$ 作为终止条件. 这种想法虽然有道理,但图 2-5表明,对于$f:\mathbf{R}^1\to\mathbf{R}^1$ 仍有这种情形,尽管$|f_{k+1}-f_k|$很小,但$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|$较大,且 $\boldsymbol{x}_k$ 远离 $\boldsymbol{x}^*$. 由此看来,将上述两种标准结合起来才是比较合理的,1972 年 D. M. Himmelblau 提出如下的收敛准则:

设 $\boldsymbol{x}_k,\boldsymbol{x}_{k+1},f_k,f_{k+1}$已经算出,$\varepsilon_1,\varepsilon_2>0$ 为预先给定的精度. 若$\|\boldsymbol{x}_k\|\leqslant\varepsilon_1$ 且$|f_k|\leqslant\varepsilon_1$(不能保证非零)时,取

$$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|<\varepsilon_2 \text{ 和 } |f_{k+1}-f_k|<\varepsilon_2 \tag{2-20}$$

作为迭代终止条件.

若$\|\boldsymbol{x}_k\|\geqslant\varepsilon_1$,且$|f_k|\geqslant\varepsilon_1$ 时,取

$$\frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|}{\|\boldsymbol{x}_k\|}<\varepsilon_2 \text{ 和} \frac{|f_{k+1}-f_k|}{|f_k|}<\varepsilon_2 \tag{2-21}$$

作为迭代终止条件.

当f_k 的值及 $\boldsymbol{x}_k$ 的分量值较大时,为减少计算,采用(2-21)作为迭代终止条件.

终止条件(2-20)和(2-21)对于解析算法及直接算法都能适用,它在算法步骤中的详细描述将在第 3 章第 3.2 节中介绍.

如果$f(\boldsymbol{x})$在 $\boldsymbol{x}^*$ 的收敛范围内是可微的,则极小点 $\boldsymbol{x}^*$ 必是稳定点,故采用$\|\nabla f(\boldsymbol{x}_k)\|<\varepsilon$ 也是合理的. 但由于鞍点也满足必要条件,因此不单独使用它.

将(2-20)或(2-21)与$\|\nabla f(\boldsymbol{x}_k)\|<\varepsilon_3$ 联合起来作为终止条件更为可靠,通常 $\varepsilon_1,\varepsilon_2$ 取 10^{-2},而 ε_3 取为 10^{-4}.

算法 2.3(H-终止法则,子程序)

设已算出 $\boldsymbol{x}_k,\boldsymbol{x}_{k+1},f_k,f_{k+1}$:

① 给定 $\varepsilon_1,\varepsilon_2,\varepsilon_3$;

② 检查$\|\nabla f(\boldsymbol{x}_{k+1})\|<\varepsilon_3$?

　是,转③;

　否,返回主程序转入下一次迭代;

③ 令 $p=\|\boldsymbol{x}_k\|$;

④ 检查 $p<\varepsilon_1$?

　是,置 $p=1$,转⑤;

⑤ 令 $p\Leftarrow\varepsilon_2 p$;

⑥ 检查$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|<p$?

　是,转⑦;

　否,返回主程序转入下一次迭代;

⑦ 令 $p=|f_k|$；

⑧ 检查 $p<\varepsilon_1$？

是，置 $p=1$，转⑨；

⑨ 令 $p\Leftarrow\varepsilon_2 p$；

⑩ 检查 $|f_{k+1}-f_k|<p$？

是，输出 x_{k+1}, f_{k+1}，迭代终止；

否，返回主程序转入下一次迭代.

其框图见图 2－6.

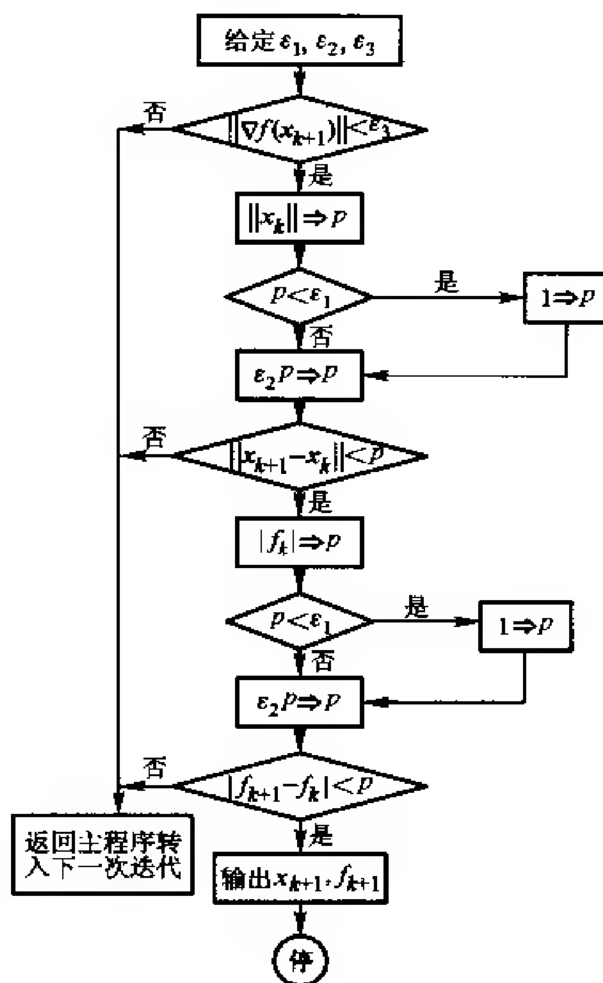


图 2－6

上述 H－终止法则仅为判定迭代点列的收敛性提供了事后检验的方法，对于一个搜索算法，即使满足下降性质，还不足以保证 $\{x_k\}$ 收敛于 $f(x)$ 的极小点，甚至不能保证 $\{x_k\}$ 收敛到 $f(x)$ 的一个稳定点. 因此必须对所给算法作收敛条件的分析.

2.3.3 收敛性条件

本节以梯度法为例，分析下降算法应具备的收敛条件. 下面的定理表明，对于梯度法，无论是完备算法还是可接受点算法，其收敛条件是相同的.

定理 2.2 设 $f:S\subset \mathbf{R}^n\to\mathbf{R}^1$，满足

(1) $S_0\in S$ 为凸集；

(2) $f(\boldsymbol{x})$ 在 S_0 上可微；

(3) 对 $\forall \boldsymbol{x},\boldsymbol{y}\in S_0$，有

$$\|\nabla f(\boldsymbol{y})-\nabla f(\boldsymbol{x})\|\leqslant L\|\boldsymbol{y}-\boldsymbol{x}\|\ (L>0);\qquad (2-22)$$

(4) 对给定的初始近似 $\boldsymbol{x}_0\in S_0$，水平集 $\phi=\{\boldsymbol{x}\mid f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_0)\}\subset S_0$，且有界；

(5) 点列 $\{\boldsymbol{x}_k\}$ 由梯度法的完备算法或可接受点算法产生，则

① $\{f(\boldsymbol{x}_k)\}$ 为单调下降，下有界序列；

② 序列 $\{\boldsymbol{x}_k\}$ 的任一极限点 $\boldsymbol{x}^*$ 满足 $\nabla f(\boldsymbol{x}^*)=0$；

(6) 若 $f(\boldsymbol{x})$ 为 S_0 上的凸函数，则 $\boldsymbol{x}^*$ 为 $f(\boldsymbol{x})$ 在 S_0 上的全局极小点.

本定理的前五个条件只能保证梯度法所产生的点列 $\{\boldsymbol{x}_k\}$ 存在极限点，且为 $f(\boldsymbol{x})$ 的稳定点，但不一定能保证 $\{\boldsymbol{x}_k\}$ 收敛(唯一极限点)，即是说 $\{\boldsymbol{x}_k\}$ 至少有一个子序列 $\{\boldsymbol{x}_{k_i}\}$ 收敛于 $f(\boldsymbol{x})$ 的稳定点. 在这种情形，即使 $\{\boldsymbol{x}_k\}$ 收敛，至多为局部收敛. 若加上 $f(\boldsymbol{x})$ 的凸性条件，不仅保证 $\{\boldsymbol{x}_k\}$ 收敛且收敛于 $f(\boldsymbol{x})$ 的全局极小点.

证 (以完备算法为例)，设 $\boldsymbol{x},\boldsymbol{x}_k\in S_0$，利用中值定理得：

$$f(\boldsymbol{x})-f(\boldsymbol{x}_k)=\nabla f(\boldsymbol{x}_k+\xi(\boldsymbol{x}-\boldsymbol{x}_k))^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}_k),\xi\in(0,1).$$

若 $\nabla f(\boldsymbol{x}_k)\neq 0$ 变换上式为：

$$f(\boldsymbol{x})-f(\boldsymbol{x}_k)=\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}_k)+(\nabla f(\boldsymbol{x}_k+\xi(\boldsymbol{x}-\boldsymbol{x}_k))-\nabla f(\boldsymbol{x}_k))^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}_k).$$

令
$$\boldsymbol{x}=\boldsymbol{x}_k-\omega\nabla f(\boldsymbol{x}_k),\quad \omega\geqslant 0.$$

$$f(\boldsymbol{x})-f(\boldsymbol{x}_k)=\omega\|\nabla f(\boldsymbol{x}_k)\|_2^2+(\nabla f(\boldsymbol{x}_k+\xi(\boldsymbol{x}-\boldsymbol{x}_k))-\nabla f(\boldsymbol{x}_k))^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}_k).$$

由 Cauchy－Schwarz 不等式，并利用(2－22)式

$$\begin{aligned}&|(\nabla f(\boldsymbol{x}_k+\xi(\boldsymbol{x}-\boldsymbol{x}_k))-\nabla f(\boldsymbol{x}_k))^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}_k)|\\ \leqslant&\|\nabla f(\boldsymbol{x}_k+\xi(\boldsymbol{x}-\boldsymbol{x}_k))-\nabla f(\boldsymbol{x}_k)\|_2\cdot\|\boldsymbol{x}-\boldsymbol{x}_k\|_2\\ \leqslant&L\|\xi(\boldsymbol{x}-\boldsymbol{x}_k)\|_2\cdot\|\boldsymbol{x}-\boldsymbol{x}_k\|_2\\ =&\xi L\omega^2\|\nabla f(\boldsymbol{x}_k)\|_2^2\end{aligned}$$

$$\leqslant L\omega^2 \| \nabla f(x_k) \|_2^2,$$

所以 $f(x) - f(x_k) \leqslant -\omega \| \nabla f(x_k) \|_2^2 + L\omega^2 \| \nabla f(x_k) \|_2^2,$

即

$$f(x_k - \omega \nabla f(x_k)) \leqslant f(x_k) - \omega \| \nabla f(x_k) \|_2^2 + L\omega^2 \| \nabla f(x_k) \|_2^2,$$

故

$$\min_{\omega} f(x_k - \omega \nabla f(x_k)) \leqslant \min_{\omega} (f(x_k) - \omega \| \nabla f(x_k) \|_2^2 + L\omega^2 \| \nabla f(x_k) \|_2^2). \tag{2-23}$$

而(2-23)中(…)内为 ω 的二次函数(且为正定),故极小点为 $\omega_k = \frac{1}{2L}$.

故由(2-23)得:

$$f(x_{k+1}) \leqslant f(x_k) - \frac{1}{4L} \| \nabla f(x_k) \|_2^2. \tag{2-24}$$

由此即知,对一切 k,序列 $\{f(x_k)\}$ 是单调下降的. 另外,由定理的假设(4),$\{x_k\} \subset \phi$ 且有界,据 Bolzano-Weierstrass 定理,必存在极限点 x^* 成立

$$\lim_{k_i \to \infty} x_{k_i} = x^*.$$

由 $f(x)$ 连续性可知,$x^* \subset \phi$ 且对一切 $k \geqslant 0$ 成立 $f(x^*) \leqslant f(x_k)$,即 $\{f(x_k)\}$ 为一单调下降序列,所以总有

$$\lim_{k_i \to \infty} (f(x_{k_{i+1}}) - f(x_{k_i})) = 0,$$

于是由(2-24)得

$$f(x_{k_{i+1}}) - f(x_{k_i}) \leqslant -\frac{1}{4L} \| \nabla f(x_{k_i}) \|_2^2,$$

从而有

$$0 = \lim_{k_i \to \infty} (f(x_{k_{i+1}}) - f(x_{k_i})) \leqslant -\frac{1}{4L} \lim_{k_i \to \infty} \| \nabla f(x_{k_i}) \|_2^2 \leqslant 0,$$

即

$$\lim_{k_i \to \infty} \| \nabla f(x_{k_i}) \|_2^2 = 0,$$

故成立

$$\nabla f(x^*) = \lim_{k_i \to \infty} \nabla f(x_{k_i}) = 0.$$

又由 $f(x)$ 在 S_0 上为凸函数,据第 1 章定理 1.12 知,$x^* \subset \phi \subset S_0$ 必为 $f(x)$ 在 S_0 上的全局极小点.

如果在定理 2.2 中,把对 $f(x)$ 一阶导数的要求加强为二阶导数的要求,即可取代 $f(x)$ 的凸性条件,不仅能够得到全局极小点的收敛性,还可给出收敛速度的估计式.

定理 2.3　设 $f: S \subset \mathbf{R}^n \to \mathbf{R}^1$,

(1) $S_0 \in S$ 为凸集;

(2) $f(\boldsymbol{x})$ 在 S_0 上二次连续可微;

(3) $f(\boldsymbol{x})$ 的 Hesse 矩阵 $H(\boldsymbol{x})$ 对 $\forall \boldsymbol{x} \in S_0$ 及非零 $\boldsymbol{y} \in \mathbf{R}^n$ 满足

$$m \| \boldsymbol{y} \|^2 \leqslant \boldsymbol{y}^{\mathrm{T}} H(\boldsymbol{x}) \boldsymbol{y} \leqslant M \| \boldsymbol{y} \|^2 \quad (M \geqslant m > 0);$$

(4) 对给定的初始近似 $\boldsymbol{x}_0 \in S_0$, $\phi = \{\boldsymbol{x} \mid f(\boldsymbol{x}) \leqslant f(\boldsymbol{x}_0)\} \subset S_0$ 有界;

(5) $\{\boldsymbol{x}_k\}$ 由梯度法的完备算法或可接受点算法产生,则

① $f(\boldsymbol{x})$ 在 S_0 上存在唯一的全局极小点 $\boldsymbol{x}^*$,且 $\boldsymbol{x}_k \to \boldsymbol{x}^*$ $(k \to \infty)$;

② 对于完备算法成立以下收敛速度估计:

$$\| \boldsymbol{x}_k - \boldsymbol{x}^* \| \leqslant c_1 q_0^{\frac{k}{2}}, c_1 < \infty,$$

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leqslant q_0^k (f(\boldsymbol{x}_0) - f(\boldsymbol{x}^*)),$$

$$q_0 = \left(1 - \frac{m^2}{M^2}\right) < 1;$$

③ 对于可接受点算法以下收敛速度估计:

$$\| \boldsymbol{x}_k - \boldsymbol{x}^* \| \leqslant c_2 q^{\frac{k}{2}}, c_2 < \infty, 0 < q < 1,$$

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leqslant q^k f(\boldsymbol{x}_0) - f(\boldsymbol{x}^*), 0 < q < 1.$$

证　因为定理 2.3 的条件包含了定理 2.2 的所有条件,所以由定理 2.2 以及第 1 章定理 1.11 即知结论①成立. 关于结论②,③即收敛速度估计的证明从略.

第 3 章　一维搜索(寻查)

如前所述,在解多元极值问题的完备算法中,每一个迭代步都需要确定 α_k 以满足

$$f(\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k)=\min_{\alpha}f(\boldsymbol{x}_k+\alpha\boldsymbol{p}_k),$$

若令 $\phi(\alpha)=f(\boldsymbol{x}_k+\alpha\boldsymbol{p}_k)$,即等价于求 α_k 满足

$$\phi(\alpha_k)=\min_{\alpha\in\mathbf{R}^1}\phi(\alpha).$$

因此,本章要讨论的问题即是求解一元极值问题

$$\min_{\alpha\in\mathbf{R}^1}\phi(\alpha),\tag{3-1}$$

其中 $\phi:\mathbf{R}^1\to\mathbf{R}^1$.

求解(3-1)式的近似方法称为一维搜索或一维寻查. 显然,通过一维寻查确定 α_k,即是在一个迭代步中把多元极值问题化为一元极值问题. 这就是一维寻查的直观含义.

下面我们进一步解释迭代点 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k$ 的空间位置. 容易证明,若从 $\boldsymbol{x}_k$ 出发,沿 $\boldsymbol{p}_k$ 方向进行一维寻查得极小点 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k$,则该点处 $f(\boldsymbol{x})$ 的梯度方向 $\nabla f(\boldsymbol{x}_{k+1})$ 与 $\boldsymbol{p}_k$ 之间满足

$$\nabla f(\boldsymbol{x}_{k+1})^{\mathrm{T}}\boldsymbol{p}_k=0.\tag{3-2}$$

事实上,$\phi'(\alpha)=\dfrac{\mathrm{d}}{\mathrm{d}\alpha}f(\boldsymbol{x}_k+\alpha\boldsymbol{p}_k)=\nabla f(\boldsymbol{x}_k+\alpha\boldsymbol{p}_k)^{\mathrm{T}}\boldsymbol{p}_k$.

由 $\qquad\phi'(\boldsymbol{x}_k)=0$,得 $\qquad\nabla f(\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k)^{\mathrm{T}}\boldsymbol{p}_k=0$,

即 $$\nabla f(\boldsymbol{x}_{k+1})^{\mathrm{T}}\boldsymbol{p}_k=0.$$

我们以二维情形为例来说明(3-2)式的几何含义. 设 AB 表示二元函数 $f(\boldsymbol{x})$ 在 x_{k+1} 处的等值线,即

$$f(\boldsymbol{x})=f(\boldsymbol{x}_{k+1}).\tag{3-3}$$

如图 3-1 所示,寻查方向 $\boldsymbol{p}_k$ 恰为等值线(3-3)在 $\boldsymbol{x}_{k+1}$ 处的切线,而极小点 $\boldsymbol{x}_{k+1}$ 恰为 $\boldsymbol{p}_k$ 相切于 AB 的切点. (3-2)式表明了 $f(\boldsymbol{x})$ 沿 $\boldsymbol{p}_k$ 方向的极小点 $\boldsymbol{x}_{k+1}$ 的平面位置.

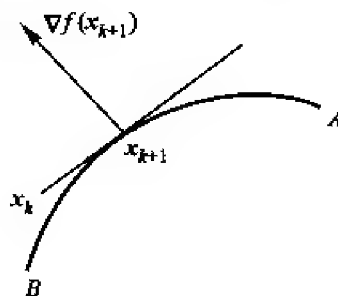


图 3-1

在微积分中解决问题(3-1)一般限于 $\phi'(\alpha)=0$ 可以直接解出的情形,在实际应用中,通常采用一维搜索方法,对 $\phi(\alpha)$ 的解

析性不作严格要求. 一维搜索方法大体可以分为三类:

1. 求根法:如二分法、Newton 法即是,它要求 $\phi(t)$ 的一阶导数或二阶导数存在且易于求出;

2. 直接法:如 0.618 法、分数法即是,它适用于一般的函数;

3. 插值法:如二次插值法、三次插值法等,它适用于一般的连续函数.

本章着重介绍二分法、0.618 法和分数法.

3.1　搜索(寻查)区间的确定

定义 3.1　设 $\phi: L\subset \mathbf{R}^1\to\mathbf{R}^1$, t^* 是 $\phi(t)$ 在 L 上的全局极小点,如果对于 L 上任意两点 t_1, t_2,且 $t_1 < t_2$,都有

$$\text{当 } t_2 \leqslant t^* \text{ 时}, \phi(t_1) > \phi(t_2);$$

$$\text{当 } t_1 \geqslant t^* \text{ 时}, \phi(t_1) < \phi(t_2).$$

则称 $\phi(t)$ 是 L 上的单谷函数(见图 3-2),显然在 L 上的严格凸函数是单谷函数,反之不必真.

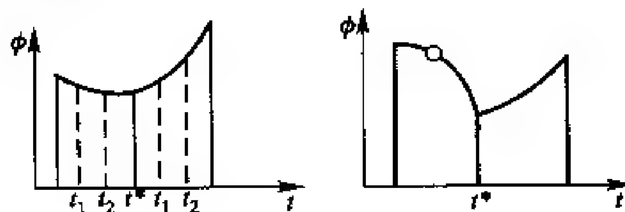


图 3-2

定义 3.2　设 $\phi: L\subset \mathbf{R}^1\to\mathbf{R}^1$ 为单谷函数, t^* 是 $\phi(t)$ 在 L 上的全局极小点,如果能找到 $t_1, t_2\in L$,使得 $t_1 < t^* < t_2$,则称 $[t_1, t_2]$ 叫做 $\phi(t)$ 的极小点 t^* 的一个寻查区间,记为 $\{t_1, t_2\}$. 若 $t_1 < t_3 < t_2$,也可把寻查区间 $\{t_1, t_2\}$ 记为 $\{t_1, t_3, t_2\}$.

定理 3.1(单谷函数的性质)　设 $\phi(t)$ 为单谷函数, $\{a, b\}$ 为其极小点的一个寻查区间,对于 $\forall t_1, t_2\in(a, b)$,且 $t_1 < t_2$,若 $\phi(t_1) < \phi(t_2)$,则 $\{a, t_2\}$ 为 $\phi(t)$ 的极小点的一个寻查区间.

若 $\phi(t_1) > \phi(t_2)$,则 $\{t_1, b\}$ 为 $\phi(t)$ 的极小点的一个寻查区间(如图 3-3).

证　设相反,在第一种情形上,假定 $\{a, t_2\}$ 不是寻查区间,则 $\phi(t)$ 的极小点 t^* 必在 $\{t_2, b\}$ 中,此时, $t_2 < t^*$,由单谷函数的定义, $\phi(t_1) >$

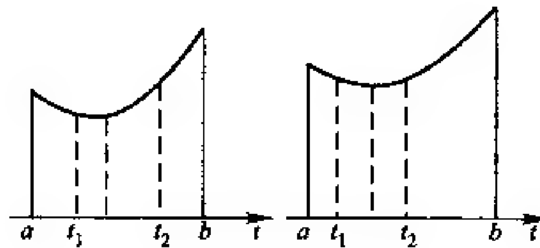


图 3-3

$\phi(t_2)$，与假设矛盾，从而$[a,t_2]$为寻查区间.

同理可证明第二种情形的论断. 根据单谷函数的上述性质，可以设计确定 $\phi(t)$极小点的寻查区间的进退算法.

设 $\phi(t)$为单谷函数，给定初始点 t_0，初始步长 h，试确定寻查区间$[a,b]$.

首先计算 $\phi(t_0)$，$\phi(t_0+h)$.

1. 前进运算：若 $\phi(t_0)>\phi(t_0+h)$，步长加大 2 倍，计算 $\phi(t_0+3h)$，再比较 $\phi(t_0+h)$与 $\phi(t_0+3h)$；

若 $\phi(t_0+h)\leqslant\phi(t_0+3h)$，则令 $a=t_0, b=t_0+3h$；否则步长加倍，继续前进.

2. 后退运算：若 $\phi(t_0)<\phi(t_0+h)$，步长缩小为$\frac{1}{4}$倍变号，计算$\phi\left(t_0-\frac{h}{4}\right)$，再比较 $\phi(t_0)$与 $\phi\left(t_0-\frac{h}{4}\right)$.

若 $\phi\left(t_0-\frac{h}{4}\right)>\phi(t_0)$，则令 $a=t_0-\frac{h}{4}, b=t_0+h$；否则步长加倍，继续后退.

后退算法的特点是：若 $\phi(t_0+ih)$下降，则前进，且每前进一步，步长加倍，直至函数值回升为止；若 $\phi(t_0+ih)$上升就后退，且每退一步，步长加倍，直到函数值回升为止.

算法 3.1(进退算法)

① 给定 t_0, h；

② 令 $a_1=t_0, a_2=t_0+h$；

③ 计算 $\phi_1=\phi(a_1)$，$\phi_2=\phi(a_2)$；

④ 检查 $\phi_1>\phi_2$?

是，转⑤(步长加倍前进)，

否，转⑦(缩小步长后退)；

⑤ 置 $h\Leftarrow 2h$，$a_2\Leftarrow a_2+h$，

$\phi_1 \Leftarrow \phi_2, \phi_2 = \phi(a_2)$;

⑥ 比较 $\phi_1 > \phi_2$?

是,令 $a_1 = a_2 - h$,转⑤,

否,转⑩

⑦ 置 $h \Leftarrow -\frac{h}{4}$;

⑧ 置 $a_1 \Leftarrow a_1 + h, \phi_2 \Leftarrow \phi_1, \phi_1 = \phi(a_1)$;

⑨ 比较 $\phi_1 > \phi_2$?

是,转⑩,

否,令 $a_2 = a_1 - h, h \Leftarrow 2h$,转⑧;

⑩ 置 $a = a_1, b = a_2$,停.

3.2 二 分 法

设 $\phi(t)$ 在寻查区间 $[a,b]$ 内具有连续的一阶导数,它的表达式为

$$\phi(t) = f(x_k + tp_k).$$

如前所述,如果 $\phi(t)$ 在 $[a,b]$ 上为单谷函数,则问题 $\min\limits_{t\in[a,b]} \phi(t)$ 可以转化为在 $[a,b]$ 上求解方程

$$\phi'(t) = 0. \tag{3-4}$$

因为 $\phi(t)$ 在 $[a,t^*]$ 和 $[t^*,b]$ 内分别是下降函数和上升函数(图 3-4),所以不妨假定

$$\phi'(a) < 0, \quad \phi'(b) > 0,$$

此时可用二分法求(3-4)式的根.

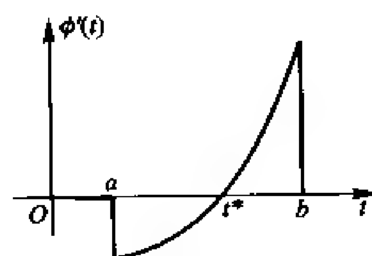


图 3-4

算法 3.2

已知 $\phi(t)$ 及 $\phi'(t)$ 的表达式,误差限 ε.

① 据算法 3.1 确定初始寻查区间 $[a,b]$,并假定 $\phi'(a) < 0, \phi'(b) > 0$;

② 计算 $c = \frac{1}{2}(a+b)$;

③ 检查 $\phi'(c) = 0$?

是,则 $t^* = c$,转⑥

④ 检查 $\phi'(c) < 0$?

是,令 $a = c$,转⑤,

否，令 $b=c$，转⑤，

⑤ 检查 $|a-b|<\varepsilon$?

是，令 $t^*=\frac{1}{2}(a+b)$，转⑥，

否，转②；

⑥ 打印 t^*，停机.

注 算法 3.2 的第一步确定 $\phi(t)$ 的寻查区间即是确定 $\phi'(t)=0$ 的有根区间 $[a,b]$. 原则上可采用进退算法 3.1，但不必照搬. 确定 $[a,b]$ 的简单作法是：

给定 $t_0, h>0$，

① 计算 $\phi'(t_0)$，若 $\phi'(t_0)=0$，令 $t^*=t_0$，停止计算；
若 $\phi'(t_0)<0$，做②，③两步；
若 $\phi'(t_0)>0$，做④，⑤两步.

② 令 $t_1=t_0+h$；

③ 计算 $\phi'(t_1)$，若 $\phi'(t_1)=0$，则令 $t^*=t_1$；
若 $\phi'(t_1)>0$，则令 $a=t_0, b=t_1$；
若 $\phi'(t_1)<0$，则 $t_0<t_1$，转②；

④ 令 $t_1=t_0-h$；

⑤ 计算 $\phi'(t_1)$，若 $\phi'(t_1)=0$，则令 $t^*=t_1$；
若 $\phi'(t_1)<0$，则令 $a=t_1, b=t_0$；
若 $\phi'(t_1)>0$，则 $t_0 \Leftarrow t_1$，转④.

二分法由于每次迭代计算量较少，而且总能收敛于一个局部极小点而受到欢迎，但它的一个突出缺陷是收敛速度很慢.

3.3 直接方法

所谓直接方法(不用求导数的方法)就是直接比较函数的大小来寻求极小点，在介绍各种直接方法之前，首先介绍如何缩短寻查区间.

假定寻查区间 $[a,b]$ 已经确定，且在 $[a,b]$ 内 $\phi(t)$ 是单谷函数，若在 $[a,b]$ 内任取二点 $t_1<t_2$，据单谷函数的性质(定理 3.1)：

若 $\phi(t_1)<\phi(t_2)$，则 $t^*\in[a,t_2]=[a_1,b_1]$；

若 $\phi(t_1)>\phi(t_2)$，则 $t^*\in[t_1,b]=[a_1,b_1]$；

若 $\phi(t_1)=\phi(t_2)$，则 $t^*\in[t_1,t_2]=[a_1,b_1]$；

继续上述过程，我们得到一串收缩区间：

$$[a,b]\to[a_1,b_1]\to\cdots\to[a_k,b_k]$$

以及每一个区间的一对比较点

$$t_1^{(k)}, t_2^{(k)} \qquad (k = 0, 1, \cdots).$$

若 $b_k - a_k < \varepsilon$，则取 $t^* = \frac{1}{2}(a_k + b_k)$.

剩下的问题是，在收缩区间内，如何选择比较点. 在上述确定 $[a_k, b_k]$ 的三种情形中，前两种情形的共同点是，每次收缩区间时，新的寻查区间内已保留了一个旧点，因此只需再选一个新点，而第三种情形，则需重选两个新点，为此，我们把三种情形合并成两种情形：

若 $\phi(t_1) < \phi(t_2)$，则 $t^* \in [a, t_2] = [a_1, b_1]$；

若 $\phi(t_1) \geqslant \phi(t_2)$，则 $t^* \in [t_1, b] = [a_1, b_1]$，这样处理就减少了每次迭代的计算量.

为了得到计算 $t_1^{(k)}, t_2^{(k)}$ 的公式，我们首先引入区间缩短率的概念.

定义 3.3(区间缩短率) 在逐次缩短区间时，设

$$\frac{b_1 - a_1}{b - a} = \tau_1 \qquad (0 < \tau_1 < 1),$$

$$\frac{b_2 - a_2}{b_1 - a_1} = \tau_2 \qquad (0 < \tau_2 < 1),$$

$$\cdots\cdots\cdots\cdots$$

$$\frac{b_k - a_k}{b_{k-1} - a_{k-1}} = \tau_k \qquad (0 < \tau_k < 1),$$

称 $\tau_k (k = 1, 2, \cdots)$ 为区间缩短率，对于 τ_k 不外乎两种情形，或者 $\tau_k =$ 常数，或者 $\tau_k \neq$ 常数，第一种情形即可引入 0.618 法，第二种情形即可引入分数法.

3.3.1 0.618 法(黄金分割法)

在定义 3.3 中，令 $\tau_k = \tau$(常数)，此时

$$\frac{b_1 - a_1}{b - a} = \frac{b_2 - a_2}{b_1 - a_1} = \cdots = \frac{b_k - a_k}{b_{k-1} - a_{k-1}} = \tau,$$

即

$$b_k - a_k = \tau^k (b - a), k = 1, 2, \cdots. \qquad (3-5)$$

根据(3-5)式，确定每次收缩后各个寻查区间内试验点 $t_1^{(k)}, t_2^{(k)} (k = 0, 1, \cdots)$ 的位置.

设在 $[a, b]$ 内 $t_1^{(0)}, t_2^{(0)}$ 的位置是对称的，见图 3-5，下面推导计算 $t_1^{(0)}, t_2^{(0)}$ 的公式.

若 $\phi(t_1^{(0)}) < \phi(t_2^{(0)})$，则应取 $a_1 = a, b_1 = t_2^{(0)}$，此时

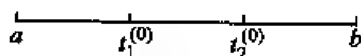


图 3-5

$$\begin{cases} t_1^{(0)} = b - (b_1 - a_1) = b - \tau(b-a), \\ t_2^{(0)} = a + (b_1 - a_1) = a + \tau(b-a). \end{cases}$$

若 $\phi(t_1^{(0)}) \geqslant \phi(t_2^{(0)})$，则应取 $a_1 = t_1^{(0)}, b_1 = b$，此时

$$\begin{cases} t_1^{(0)} = b - (b_1 - a_1) = b - \tau(b-a), \\ t_2^{(0)} = a + (b_1 - a_1) = a + \tau(b-a). \end{cases}$$

因此，对于上述两种情形，确定 $t_1^{(0)}, t_2^{(0)}$ 的公式是一致的，即

$$\begin{cases} t_1^{(0)} = b - \tau(b-a) = a + (1-\tau)(b-a), \\ t_2^{(0)} = a + \tau(b-a). \end{cases} \tag{3-6}$$

一般地，假定 $t_1^{(k)}, t_2^{(k)}$ 在 $[a_k, b_k]$ 内的位置是对称的，则应有

$$\begin{cases} t_1^{(k)} = b_k - \tau(b_{k+1} - a_{k+1}) = a_k + (1-\tau)(b_k - a_k), \\ t_2^{(k)} = a_k + \tau(b_k - a_k). \end{cases} \tag{3-7}$$

(3-7)式就是 0.618 法确定试验点的计算公式. 它表明，在寻查区间 $[a_k, b_k]$ $(k=0,1,\cdots)$ 内，试验点 $t_1^{(k)}$ 位于区间长的 $1-\tau$ 倍处，而另一试验点 $t_2^{(k)}$ 位于区间长的 τ 倍处，后面将进一步导出 $\tau = 0.618$.

算法 3.3

已知 $\phi(t)$，误差限 ε，$\tau = 0.618$.

① 确定 $\phi(t)$ 的初始寻查区间 $[a,b]$（算法 3.1）；

② 计算 $t_1 = a + (1-\tau)(b-a)$，$\phi_1 = \phi(t_1)$；

③ 计算 $t_2 = a + \tau(b-a)$，$\phi_2 = \phi(t_2)$；

④ 检查 $|t_1 - t_2| < \varepsilon$?

是，令 $t^* = \frac{1}{2}(t_1 + t_2)$，打印 t^*，停止计算；

⑤ 检查 $\phi_1 < \phi_2$?

是，置 $b = t_2, t_2 = t_1, \phi_2 = \phi_1, t_1 = a + (1-\tau)(b-a), \phi_1 = \phi(t_1)$，转④；

⑥ 置 $a = t_1, t_1 = t_2, \phi_1 = \phi_2, t_2 = a + \tau(b-a), \phi_2 = \phi(t_2)$，转④.

在算法 3.3 中，当收缩区间为 $[a,b]$ 时，在新的寻查区间 $[a_1, b_1]$ 内，确定试验点，$t_1^{(1)}, t_2^{(1)}$ 的方法是：

当 $\phi_1<\phi_2$ 时,保留旧点 $t_1^{(0)}\Rightarrow t_2^{(1)}$,取新点

$$t_1^{(1)}=a_1+(1-\tau)(b_1-a_1);$$

当 $\phi_1\geqslant\phi_2$ 时,保留旧点 $t_2^{(0)}\Rightarrow t_1^{(1)}$,取新点

$$t_2^{(1)}=a_1+\tau(b_1-a_1).$$

在第一种情形,$t_2^{(1)}=t_1^{(0)}$,此时有

$$a_1+\tau(b_1-a_1)=a+(1-\tau)(b-a);$$

而 $a=a_1,b_1-a_1=\tau(b-a)$,

故得
$$\tau^2+\tau-1=0.$$

因为 $0<\tau<1$,故 $\tau=\dfrac{-1+\sqrt{5}}{2}\approx0.618.$

在第二种情形,$t_1^{(1)}=t_2^{(0)}$,此时有

$$a_1+(1-\tau)(b_1-a_1)=a+\tau(b-a);$$

而 $a_1=t_1^0=a+(1-\tau)(b-a)$, $b_1-a_1=\tau(b-a)$,

故仍得
$$\tau^2+\tau-1=0.$$

由 $0<\tau<1$,故 $\tau\approx0.618$.

3.3.2　分数法(Fibonacci 法)

在定义 3.3 中,若 $\tau_k\neq$ 常数,取

$$\tau_k=\frac{F_{k-1}}{F_k}$$

其中 F_k 满足递推式

$$\begin{cases}F_0=F_1=1,\\ F_k=F_{k-1}+F_{k-2},\end{cases}\quad k=2,3,\cdots,\tag{3-8}$$

即

k	0	1	2	3	4	5	6	…
F_k	1	1	2	3	5	8	13	…

由(3-8)所确定的数列叫做 Fibonacci 数. 下面给出分法的计算步骤:

设 $\phi(t)$ 在 $[a,b]$ 上是单谷函数,且给定试验点的个数为 n,即用分数法一共要计算 n 次函数值,于是可列出 Fibonacci 数 $F_0,F_1,\cdots,F_n$.

1. 第一次收缩将区间 $[a,b]$ 缩小为 $[a_1,b_1]$.

首先在 $[a,b]$ 上按如下公式计算 $t_1^{(0)},t_2^{(0)}$:

令
$$\begin{cases}\tau_n=\dfrac{F_{n-1}}{F_n},\\[2ex] t_1^{(0)}=b-\tau_n(b-a)=b-\dfrac{F_{n-1}}{F_n}(b-a)=a+\dfrac{F_{n-2}}{F_n}(b-a),^{①}\\[2ex] t_2^{(0)}=a+\tau_n(b-a)=a+\dfrac{F_{n-1}}{F_n}(b-a).\end{cases}\qquad(3-9)$$

然后比较 $\phi(t_1^{(0)}),\phi(t_2^{(0)})$，确定 $[a_1,b_1]$：

若 $\phi(t_1^{(0)})<\phi(t_2^{(0)})$，取 $a_1=a,b_1=t_2^{(0)}$；

若 $\phi(t_1^{(0)})\geqslant\phi(t_2^{(0)})$，取 $a_1=t_1^{(0)},b_1=b$.

2. 第二次收缩将区间 $[a_1,b_1]$ 缩小为 $[a_2,b_2]$.

首先在 $[a_1,b_1]$ 内按如下方式确定试验点 $t_1^{(1)},t_2^{(1)}$：

若 $\phi(t_1^{(0)})<\phi(t_2^{(0)})$，取
$$\begin{cases}t_2^{(1)}=t_1^{(0)}\quad\text{（保留旧点）},\\[1ex] t_1^{(1)}=a_1+\dfrac{F_{n-3}}{F_{n-1}}(b_1-a_1);\end{cases}\qquad(3-10)$$

若 $\phi(t_1^{(0)})\geqslant\phi(t_2^{(0)})$，取
$$\begin{cases}t_1^{(1)}=t_2^{(0)}\quad\text{（保留旧点）},\\[1ex] t_2^{(1)}=a_1+\dfrac{F_{n-2}}{F_{n-1}}(b_1-a_1);\end{cases}\qquad(3-11)$$

然后比较 $\phi(t_1^{(1)}),\phi(t_2^{(1)})$，确定 $[a_2,b_2]$：

若 $\phi(t_1^{(1)})<\phi(t_2^{(1)})$，取 $a_2=a_1,b_2=t_2^{(1)}$；

若 $\phi(t_1^{(1)})\geqslant\phi(t_2^{(1)})$，取 $a_2=t_1^{(1)},b_2=b_1$.

可以证明在 $[a_1,b_1]$ 内用类似于(3－9)的公式确定 $t_1^{(1)},t_2^{(1)}$ 与(3－10)、(3－11)完全一致.事实上，按(3－9)式，
$$t_1^{(1)}=b_1-\tau_{n-1}(b_1-a_1)=b_1-\frac{F_{n-2}}{F_{n-1}}(b_1-a_1)=a_1+\frac{F_{n-3}}{F_{n-1}}(b_1-a_1),$$
$$t_2^{(1)}=a_1-\tau_{n-1}(b_1-a_1)=a_1-\frac{F_{n-2}}{F_{n-1}}(b_1-a_1).$$

若 $\phi(t_1^{(0)})<\phi(t_2^{(0)})$，此时由于 $a_1=a,b_1=t_2^{(0)}$，故有

---

① $$b-\frac{F_{n-1}}{F_n}(b-a)=a+(b-a)-\frac{F_{n-1}}{F_n}(b-a)=a+\left(1-\frac{F_{n-1}}{F_n}\right)(b-a)$$
$$=a+\frac{F_n-F_{n-1}}{F_n}(b-a)=a+\frac{F_{n-2}}{F_n}(b-a).$$

$$t_2^{(1)} = a + \frac{F_{n-2}}{F_{n-1}}\left(a + \frac{F_{n-1}}{F_n}(b-a) - a\right) = a + \frac{F_{n-2}}{F_n}(b-a) = t_1^{(0)},$$

由此得

$$\begin{cases} t_2^{(1)} = t_1^{(0)}, \\ t_1^{(1)} = a_1 + \dfrac{F_{n-3}}{F_{n-1}}(b_1 - a_1), \end{cases}$$

此即(3－10).

若 $\phi(t_1^{(0)}) > \phi(t_2^{(0)})$，此时由于 $a_1 = t_1^{(0)}, b_1 = b$，故有

$$t_1^{(1)} = t_1^{(0)} + \frac{F_{n-3}}{F_{n-1}}(b - t_1^{(0)}) = a + \frac{F_{n-1}}{F_n}(b-a) = t_2^{(0)},$$

由此得

$$\begin{cases} t_1^{(1)} = t_2^{(0)}, \\ t_2^{(1)} = a_1 + \dfrac{F_{n-2}}{F_{n-1}}(b_1 - a_1), \end{cases}$$

此即(3－11).

3. 设已进行 $n-1$ 次区间收缩，即将 $[a_{n-2}, b_{n-2}]$ 缩小为 $[a_{n-1}, b_{n-1}]$. 在 $[a_{n-2}, b_{n-2}]$ 内按如下公式确定试验点：

$$\begin{cases} t_1^{(n-2)} = b_{n-2} - \dfrac{F_1}{F_2}(b_{n-2} - a_{n-2}) = a_{n-2} + \dfrac{F_0}{F_2}(b_{n-2} - a_{n-2}), \\ t_2^{(n-2)} = a_{n-2} + \dfrac{F_1}{F_2}(b_{n-2} - a_{n-2}). \end{cases} \tag{3－12}$$

因为 $F_0 = F_1$，故 $t_1^{(n-2)} = t_2^{(n-2)}$，由此不能比较 $\phi(t_1^{(n-2)})$，$\phi(t_2^{(n-2)})$ 的大小，将(3－12)改写成：

$$\begin{cases} t_1^{(n-2)} = a_{n-2} - \dfrac{1-\delta}{2}(b_{n-2} - a_{n-2}), \\ t_2^{(n-2)} = a_{n-2} + \dfrac{1+\delta}{2}(b_{n-2} - a_{n-2}). \end{cases} \tag{3－13}$$

式中 δ 为任意小的正数. 然后比较 $\phi(t_1^{(n-2)})$，$\phi(t_2^{(n-2)})$：

若 $\phi(t_1^{(n-2)}) < \phi(t_2^{(n-2)})$，取 $t^* = t_1^{(n-2)}$；

若 $\phi(t_1^{(n-2)}) \geqslant \phi(t_2^{(n-2)})$，取 $t^* = t_2^{(n-2)}$.

从上述步骤中可以看出：分数法与0.618法的迭代过程基本上是相同的，所不同的只有两点：一是区间缩短率 $\tau_k = F_{k-1}/F_k$ 并非常数；另一点是要事先确定试验点的个数 n.

如何定 n? 我们作如下分析：

在收缩区间过程中有

$$[a,b]\xrightarrow{\tau_n=\frac{F_{n-1}}{F_n}}[a_1,b_1]\xrightarrow{\tau_{n-1}=\frac{F_{n-2}}{F_{n-1}}}[a_2,b_2]\longrightarrow\cdots$$

$$\xrightarrow{\tau_3=\frac{F_2}{F_3}}[a_{n-2},b_{n-2}]\xrightarrow{\tau_2=\frac{F_1}{F_2}}[a_{n-1},b_{n-1}].$$

按定义 3.3,有

$$\frac{b_1-a_1}{b-a}=\frac{F_{n-1}}{F_n}=\tau_n,$$

$$\frac{b_2-a_2}{b_1-a_1}=\frac{F_{n-2}}{F_{n-1}}=\tau_{n-1},$$

$$\cdots\cdots\cdots\cdots$$

$$\frac{b_{n-2}-a_{n-2}}{b_{n-3}-a_{n-3}}=\frac{F_2}{F_3}=\tau_3,$$

$$\frac{b_{n-1}-a_{n-1}}{b_{n-2}-a_{n-2}}=\frac{F_1}{F_2}=\tau_2,$$

还有

$$\frac{b_1-a_1}{b-a}\cdot\frac{b_2-a_2}{b_1-a_1}\cdot\cdots\cdot\frac{b_{n-2}-a_{n-2}}{b_{n-3}-a_{n-3}}\cdot\frac{b_{n-1}-a_{n-1}}{b_{n-2}-a_{n-2}}$$

$$=\frac{F_{n-1}}{F_n}\cdot\frac{F_{n-2}}{F_{n-1}}\cdot\cdots\cdot\frac{F_2}{F_3}\cdot\frac{F_1}{F_2},$$

即
$$\frac{b_{n-1}-a_{n-1}}{b-a}=\frac{F_1}{F_n}=\frac{1}{F_n},$$

亦即
$$b_{n-2}-a_{n-2}=\frac{1}{F_n}(b-a).$$

若欲使最后区间长度小于某个正数 $\varepsilon>0$,即

$$b_{n-2}-a_{n-2}<\varepsilon,$$

则只要选择 n,使

$$F_n>\frac{1}{\varepsilon}(b-a),$$

于是便可得到对应的 n.

最后我们给出分数法的计算方法.

算法 3.4

① 给定初始寻查区间 $[a,b]$,及 $\varepsilon>0,\delta>0$;

② 令 $F=\frac{1}{\varepsilon}(b-a)$;$n=1,F_0=1,F_1=1$;

③ 置 $n=n+1$;$F_n=F_{n-2}+F_{n-1}$;

④ 若 $F_n \geq F$，转⑤，否则转③；

⑤ 置 $k=0$，

$$t_1 = a + \frac{F_{n-2}}{F_n}(b-a),$$

$$t_2 = a + \frac{F_{n-1}}{F_n}(b-a),$$

并计算 $\phi_1 = \phi(t_1)$，$\phi_2 = \phi(t_2)$；

⑥ 若 $\phi_1 < \phi_2$，则令 $a=a, b=t_2; t_2=t_1, \phi_2=\phi_1$；

$$t_1 = a + \frac{F_{n-k-3}}{F_{n-k-1}}(b-a), \phi_1 = \phi(t_1);\text{转 ⑦};$$

若 $\phi_1 \geq \phi_2$，则令 $b=b, a=t_1; t_1=t_2, \phi_1=\phi_2$；

$$t_2 = a + \frac{F_{n-k-2}}{F_{n-k-1}}(b-a), \phi_2 = \phi(t_2);$$

⑦ 置 $k=k+1$，

若 $k<n-3$，则转⑥，

若 $k=n-3$，转⑧；

⑧ 令

$$t_1 = a + \frac{1-\delta}{2}(b-a), \phi_1 = \phi(t_1),$$

$$t_2 = a + \frac{1+\delta}{2}(b-a), \phi_2 = \phi(t_2);$$

⑨ 若 $\phi_1 < \phi_2$，令 $t^* = t_1$；

若 $\phi_1 \geq \phi_2$，令 $t^* = t_2$.

注　在算法 3.4 中，由于 $k=0$ 时，公式已列出

$$t_1 = a + \frac{F_{n-2}}{F_n}(b-a),$$

$$t_2 = a + \frac{F_{n-1}}{F_n}(b-a),$$

故通式应为

$$t_1 = a + \frac{F_{n-k-3}}{F_{n-k-1}}(b-a),$$

$$t_2 = a + \frac{F_{n-k-2}}{F_{n-k-1}}(b-a).$$

判别式应为 $k=n-3$.

第 4 章　Newton 方法及其改进

本章讨论 Newton 方法及改进它的几种修正策略，着重介绍执行算法所采取的各种保护措施，略去了算法收敛性分析.

4.1　Newton 方法及其局限性

Newton 方法的基本思想是用一个二次函数去近似目标函数，然后精确求出这个二次函数的极小点，以它作为目标函数极小点的近似值.

设 $f:\mathbf{R}^n\to\mathbf{R}^1$ 在其极小点 $\boldsymbol{x}^*$ 的某一开凸邻域

$$D = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| < \delta\}$$

内二次可微，将 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}_k$ 处展开到二次项

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{G}_k(\boldsymbol{x} - \boldsymbol{x}_k), \quad (4-1)$$

其中

$$\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k), \quad \boldsymbol{G}_k = \nabla^2 f(\boldsymbol{x}_k),$$

令

$$Q(\boldsymbol{x}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{G}_k(\boldsymbol{x} - \boldsymbol{x}_k), \quad (4-2)$$

当 $\boldsymbol{G}_k$ 正定时，二次函数 $Q(\boldsymbol{x})$ 有极小点，由第 1 章定理 1.10（极小点的一阶必要条件）知

$$\nabla Q(\boldsymbol{x}) = \boldsymbol{g}_k + \boldsymbol{G}_k(\boldsymbol{x} - \boldsymbol{x}_k) = 0.$$

由此得 $\boldsymbol{x} = \boldsymbol{x}_k - \boldsymbol{G}_k^{-1}\boldsymbol{g}_k$ 即是 $Q(\boldsymbol{x})$ 的极小点. 令

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{G}_k^{-1}\boldsymbol{g}_k, \quad (4-3)$$

我们用 $\boldsymbol{x}_{k+1}$ 作为 $f(\boldsymbol{x})$ 极小点 $\boldsymbol{x}^*$ 的第 $k+1$ 次近似，(4－3)中 $\boldsymbol{p} = -\boldsymbol{G}_k^{-1}\boldsymbol{g}_k$ 是 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}_k$ 处的 Newton 方向，因此称该式为 Newton 迭代公式.

以二元函数为例说明 Newton 法的几何意义. 如果 $\boldsymbol{G}_k$ 正定，则(4－2)所示的二次函数 $Q(\boldsymbol{x})$ 的等值线是一族椭圆，由(4－3)得到的 $\boldsymbol{x}_{k+1}$ 恰为这族椭圆的中心. 若以过 $\boldsymbol{x}_k$ 点的椭圆近似代替 $f(\boldsymbol{x})$ 过 $\boldsymbol{x}_k$ 点的等值线，则 $\boldsymbol{x}_{k+1}$ 可作为 $f(\boldsymbol{x})$ 的极小点 $\boldsymbol{x}^*$ 的第 $k+1$ 级近似值（见图 4－1）.

算法 4.1（Newton 法）

设 $\boldsymbol{x}_0$ 是 $f(\boldsymbol{x})$ 极小点 $\boldsymbol{x}^*$ 的初始近似.

① 置 $k=0$;

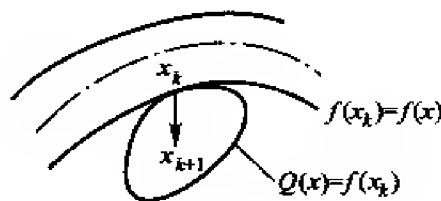


图 4-1

② 计算 $g_k=(g_1^{(k)},\cdots,g_n^{(k)})^T$ 和 $G_k=(G_{ij}^{(k)}),i,j=1,2,\cdots,n$,

其中
$$g_i^{(k)}=\frac{\partial f(x_k)}{\partial x_i^{(k)}},i=1,2,\cdots,n,$$

$$G_{ij}^{(k)}=\frac{\partial^2 f(x_k)}{\partial x_j^{(k)}\partial x_i^{(k)}},i,j=1,2,\cdots,n,$$

③ 计算迭代方向 p_k: $G_k p_k=-g_k$;

④ 计算 $x_{k+1}=x_k+p_k$;

⑤ 检查迭代终止条件(即执行第 2 章算法 2.3)

若满足,则输出 x_{k+1},停止计算,否则转⑥;

⑥ 置 $k=k+1$ 转②.

例 4.1 求 $f(x)=f(x_1,x_2)=(x_1-2)^4+(x_1-2x_2)^2$ 的极小点.

解 设初始点为 $x_0=(x_1^{(0)},x_2^{(0)})^T=(0,3)^T$,直接计算可得

$$g_0=\nabla f(x_0)=(-44,24)^T,$$

$$G_0=\nabla^2 f(x_0)=\begin{bmatrix}50 & -4\\ -4 & 8\end{bmatrix},$$

解方程
$$G_0p_0=-g_0,$$
得
$$p_0=(0.67,-2.67)^T,$$
于是
$$x_1=x_0+p_0=(0.67,0.33)^T.$$
至此,第一次迭代完毕.按照同样的步骤,只要 $g_k\neq 0$,迭代就可以继续下去,下面列出前六次迭代的结果(见表 4-1).

表 4-1

k	0	1	2	3	4	5	6
x_k	$\begin{bmatrix}0.00\\3.00\end{bmatrix}$	$\begin{bmatrix}0.67\\0.33\end{bmatrix}$	$\begin{bmatrix}1.11\\0.56\end{bmatrix}$	$\begin{bmatrix}1.41\\0.70\end{bmatrix}$	$\begin{bmatrix}1.61\\0.80\end{bmatrix}$	$\begin{bmatrix}1.74\\0.87\end{bmatrix}$	$\begin{bmatrix}1.83\\0.91\end{bmatrix}$
$f(x_k)$	52	3.13	0.63	0.12	0.02	0.005	0.000 9

显然迭代点列 $\{x_k\}$ 当 k 充分大时,趋于 $f(x)$ 的极小点 $x^*=(2,1)^T$.

当目标函数 $f(x)$ 为正定二次函数时,(4-1)式为精确等式:即 $f(x)$

$=Q(x)$,用(4-3)式进行一次迭代就能达到极小点.

在一般情形,可以证明,如果初始点 x_0 充分接近 x^*,在一定条件下,则由算法4.1所产生的点列 $\{x_k\}$ 以二阶收敛速度收敛于 $f(x)$ 的稳定点(驻点).若对 $f(x)$ 加上凸性的假定,则 x^* 为 $f(x)$ 的局部(或全局)极小点.但由于Newton法的简单算法对初始点 x_0 的依赖很大,因而实际应用是不合适的.在实际应用中:通常由于 x_0 远离 x^*,从而使Newton算法可能出现一些异常情况(如考虑第 k 次迭代的情形)如下:

(1) G_k^{-1} 不存在,导致 $p_k=-G_k^{-1}g_k$ 无意义;

(2) G_k^{-1} 存在但不正定,由于 $g_k^Tp_k=-g_kG_k^{-1}g_k$ 非负,故知 p_k 非下降方向,导致 $f_{k+1}\geqslant f_k$;

(3) G_k^{-1} 存在且正定,但若 p_k 很大,因而简单算法 $x_{k+1}=x_k+p_k$ 不能保证 $f_{k+1}<f_k$;

(4) p_k 与 g_k 正交,p_k 既非下降方向亦非上升方向.

当出现上述任何一种情形时,Newton迭代法便无法继续迭代下去.

下面我们讨论为了避免出现前面所列出的一些异常情形而采取的一些保护措施,即讨论Newton算法的几种修正策略.

4.2 Newton算法的改进

1. 若出现情形(1),可改变寻查方向,采用最速下降方向替换Newton方向,即取

$$p_k=-g_k;$$

2. 若出现情形(2),即 $g_k^Tp_k\geqslant 0$,当

$$q_k^Tp_k>\varepsilon\|g_k\|\cdot\|p_k\| \qquad (\text{非正交})$$

时,显然Newton方向非下降方向,但其反方向 $p_k=-p_k=G_kg_k$ 为下降方向,故可沿此方向进行搜索;

3. 若出现情形(3),即若 p_k 过大,我们采用对步长阻尼的办法,令

$$x_{k+1}=x_k+\alpha_kp_k,$$

其中 $\alpha_k\in(0,1]$,可导致 $f_{k+1}<f_k$,α_k 称为迭代步长的阻尼系数,可由下面的算法给出.

算法4.2(子程序)

设 $x_k,p_k,f_k,\varepsilon>0$ 已给定,

① 置 $\alpha=1$;

② 计算 $x=x_k+\alpha p_k$,

$$f = f(x);$$

③ 检查 $f<f_k$?

是,令 $\alpha_k=\alpha, f_{k+1}=f$,返回主程序,计算 x_{k+1};

④ 检查 $|\alpha|<\varepsilon$?

是,停止计算(算法失败),否则转⑤;

⑤ 置 $\alpha=\dfrac{\alpha}{2}$,转②.

4. 若出现情形(4),即 $g_k^T p_k=0$,在实际计算中,当

$$|g_k^T p_k| \leqslant \varepsilon \|g_k\| \cdot \|p_k\|$$

时,即说明 p_k 几乎与 g_k 正交,显然此时的 Newton 方向 p_k 是不利方向,可改变迭代方向,令 $p_k=-g_k$.

综合上述讨论,得到下面改进的 Newton 算法.

算法 4.3

① 送 $x_0, \varepsilon>0$;

② 置 $k=0$;

③ 计算 g_k 和 G_k:

$$g_i^{(k)} = \frac{\partial f(x_k)}{\partial x_i^{(k)}},\ i=1,2,\cdots,n,$$

$$G_{ij}^{(k)} = \frac{\partial^2 f(x_k)}{\partial x_j^{(k)} \partial x_i^{(k)}},\ i,j=1,2,\cdots,n;$$

④ 检查 $\det(G_k)=0$?

是,转⑫;

⑤ 计算 $p_k=-G_k^{-1}g_k$,或解方程组 $G_k p_k=-g_k$;

⑥ 检查 $|g_k^T p_k| \leqslant \varepsilon \|g_k\| \cdot \|p_k\|$?

是,转⑫;

⑦ 检查 $g_k^T p_k > \varepsilon \|g_k\| \cdot \|p_k\|$?

是,转⑬;

⑧ 执行算法 4.2;

⑨ 置 $x_{k+1}=x_k+\alpha_k p_k$;

⑩ 执行第 2 章算法 2.3,不满足 H - 终止法则转⑪,否则转⑭;

⑪ 置 $k=k+1$,转③;

⑫ 置 $p_k=-g_k$,转⑧;

⑬ 置 $p_k=-p_k$,转⑧;

⑭ 置 $x^*=x_{k+1}$;

⑮ Stop.

4.3 特征值法(Greenstadt 方法)

由 4.1 的分析可知,Newton 法往往由于 Hesse 矩阵 $G(x)=\nabla^2 f(x)$ 不满足正定要求(或者甚至奇异)而遭到破坏,因此强迫 Hesse 矩阵正定是修正 Newton 法的一条重要途径. 下面介绍的特征值法即是利用 Hesse 矩阵的特征值系统来构造正定矩阵的一种方法.

定理 4.1(实对称矩阵的谱分解) 设 G 为实对称矩阵,$\lambda_i(i=1,\cdots,n)$ 为 G 的特征值,$u_i(i=1,\cdots,n)$ 是对应于 λ_i 的特征向量,则

$$G=\sum_{i=1}^{n}\lambda_i u_i u_i^{\mathrm{T}}.$$

证 由题设知
$$Gu_i=\lambda_i u_i \quad (i=1,\cdots,n), \tag{4-4}$$

令
$$G=\begin{bmatrix} G_1^{\mathrm{T}} \\ \vdots \\ G_n^{\mathrm{T}} \end{bmatrix}, \tag{4-5}$$

其中 $G_i^{\mathrm{T}}(i=1,\cdots,n)$ 是 G 的第 i 行,故由(4-4)和(4-5)有
$$G_j^{\mathrm{T}}u_i=\lambda_i u_{ji} \quad (i,j=1,\cdots,n), \tag{4-6}$$

其中
$$u_i=(u_{1i},\cdots,u_{ni})^{\mathrm{T}},\quad i=1,\cdots,n,$$

由于 $u_m(m=1,\cdots,n)$ 组成 $\mathbf{R}^n$ 上的一组基,于是对于每一个 j 均存在一组实数 u_{jm} $(j,m=1,\cdots,n)$,使得

$$G_j=\sum_{m=1}^{n}u_{jm}u_m. \tag{4-7}$$

由(4-6)和(4-7)

$$\sum_{m=1}^{n}u_{jm}u_m^{\mathrm{T}}u_i=\lambda_i u_{ji},$$

再由 $u_i(i=1,\cdots,n)$ 的正交性,知
$$u_{ji}=\lambda_i u_{ji} \quad (i,j=1,\cdots,n), \tag{4-8}$$

综合(4-5),(4-7)和(4-8)式得

$$G=\sum_{i=1}^{n}\lambda_i u_i u_i^{\mathrm{T}}.$$

推论 4.1 设 G 为实对称阵,且 $\det(G)\neq 0$,则

$$G^{-1}=\sum_{i=1}^{n}\lambda_i^{-1}u_i u_i^{\mathrm{T}}. \tag{4-9}$$

证 由于 G 为实对称阵,且 G^{-1} 亦为对称阵,且特征值系统为

λ_i^{-1} $(i=1,\cdots,n)$ 是 G^{-1} 的特征值,

$u_i\ (i=1,\cdots,n)$ 是对应于 λ_i^{-1} 的特征向量.

故由定理 4.1 知

$$G^{-1}=\sum_{i=1}^{n}\lambda_i^{-1}u_iu_i^{-1}.$$

根据上述定理及推论,我们可以就一个实对称阵来构造与之相应的正定阵.

设 G 为非奇异的实对称阵,其特征值系统是

$\lambda_i\neq 0\ (i=1,\cdots,n)$ 是 G 的特征值,

$u_i\ (i=1,\cdots,n)$ 为对应于 λ_i 的特征向量,

则
$$\hat{G}=\sum_{i=1}^{n}|\lambda_i|u_iu_i^{\mathrm{T}}$$

为正定矩阵,且当 G 正定时,$\hat{G}=G$.

设 G 为实对称阵,令 $\delta>0$ 为一正小数,设

$$\gamma_i=\max\{|\lambda_i|,\delta\},\ i=1,2,\cdots,n,$$

则
$$\overline{G}=\sum_{i=1}^{n}\lambda_iu_iu_i^{\mathrm{T}}$$

为正定阵,且

$$\overline{G}^{-1}=\sum_{i=1}^{n}\gamma_i^{-1}u_iu_i^{\mathrm{T}}$$

亦为正定阵,显然,当 G 正定时,则 G,G^{-1}不仅正定,且

$$\overline{G}=G,\quad \overline{G}^{-1}=G^{-1}.$$

根据以上讨论,形成 Greenstadt 特征值法.

算法 4.4

① 送 $x_0,\delta>0$;

② 置 $k=0$;

③ 计算 g_k 和 G_k;

④ 用幂法计算 G_k 的特征值 $\lambda_j^{(k)}$ 和特征向量 $u_j^{(k)}\quad(j=1,\cdots,n)$;

⑤ 令 $u_j^{(k)}=\dfrac{u_j^{(k)}}{\|u_j^{(k)}\|}\qquad(j=1,\cdots,n)$;

⑥ 令 $\gamma_i=\max\{|\lambda_i|,\delta\}\qquad(i=1,\cdots,n)$;

⑦ 计算 $\overline{G}_k^{-1}=\sum_{i=1}^{n}\gamma_i^{-1}u_i^{(k)}u_i^{(k)\mathrm{T}}$;

⑧ 计算 $\overline{p}_k=-\overline{G}_k^{-1}g_k$,成解方程组

$$\overline{G}_k\overline{p}_k=-g_k;$$

⑨ 执行算法 4.2 后返回主程序⑩;

⑩ 计算 $x_{k+1}=x_k+\alpha_k p_k$;

⑪ 执行第 1 章算法 1.3,或者转⑫,或者转⑬;

⑫ 置 $k=k+1$;

⑬ 置 $x^*=x_{k+1}$;

⑭ Stop.

4.4 Newton 算法的 Gill 和 Murray 修正方案

特征值法由于在每一个迭代步利用 G_k 的特征值系统来构造正定矩阵

$$\overline{G}=\sum_{i=1}^{n}\gamma_i u_i^{(k)}u_i^{(k)\mathrm{T}},$$

因此计算量是很大的. 事实上,对于大多数问题,只是在少量迭代步骤,G_k 才不能保持正定. 1972 年 Murray 对 Greenstadt 方法作了进一步修正,即仅当 G_k 非正定时才构造正定阵 $\overline{G}_k$. 基本想法是,当 G_k 正定时直接对 G_k 实行 LL^{T} 分解;当 G_k 非正定时,用正定矩阵 $\overline{G}_k=G_k+D_k$ 代替 G_k(其中 D_k 为对角阵),并对 $\overline{G}$ 实行 LL^{T} 分解,即

$$\overline{G}_k=\overline{L}_k\overline{L}_k^{\mathrm{T}}.$$

构造 $\overline{L}_k$ 时,应注意如下几点:

(1) 不需要先给出 D_k 去直接分解 $\overline{G}_k=G_k+D_k$,而是利用 G_k 的元素及一个正小数 $\delta>0$ 去构造 $\overline{L}_k$ 的元素;

(2) 构造下三角阵 $\overline{L}_k$ 的元素 $\overline{l}_{ij}$时,需满足 $\overline{l}_{ij}>0$,从而保证 L_k 正定;

(3) 要求 $\overline{L}_k$ 的下三角部分元素 $\overline{l}_{ij}(i<j)$ 有界,从而保证算法的稳定性;

(4) 当 G_k 正定时,$\overline{G}_k=\overline{L}_k\overline{L}_k^{\mathrm{T}}=L_kL_k^{\mathrm{T}}=G_k$.

下面给出对 $\overline{G}_k$ 进行 $\overline{L}_k\overline{L}_k^{\mathrm{T}}$ 分解的算法.

算法 4.5

给定 $G=(g_{ij})(i,j=1,\cdots,n)$ 及 $\delta>0$.

一、计算 $\overline{L}$ 阵的第一列

① 计算判别数 β:

(1) 计算 $\xi_j=\max\limits_{j\le i\le n}|g_{ij}|\quad(j=1,2,\cdots,n-1)$;

(2) 计算 $\xi=\max\limits_{1\le j\le n-1}|\xi_j|$;

(3) 计算 $\gamma=\max\limits_{1\le i\le n}|g_{ii}|$;

(4) 计算 $\beta = \max\left\{\gamma^{\frac{1}{2}}, \left(\frac{\xi}{n}\right)^{\frac{1}{2}}\right\}$.

② 计算辅助阵 $\hat{L}$ 的第一列(下三角阵);

(1) 计算 $\hat{l}_1 = \max\{\delta, |g_{11}|^{\frac{1}{2}}\}$;

(2) 计算 $\hat{l}_i = \frac{g_{i1}}{\hat{l}_1}$　$(i = 2, \cdots, n)$.

③ 计算判别数 $\theta = \max\limits_{2 \leq i \leq n} |\hat{l}_i|$;

④ 计算 $\bar{l}_{i1}$　$(i = 1, \cdots, n)$;

若 $\theta \leq \beta$,令 $\bar{l}_{i1} = \hat{l}_i$,

若 $\theta > \beta$,令

$$\begin{cases} \bar{l}_{11} = \frac{\theta}{\beta}\hat{l}_1, \\ \bar{l}_{i1} = \frac{\beta}{\theta}\hat{l}_i, \end{cases} \quad (i = 2, \cdots, n).$$

二、计算 $\bar{L}$ 阵的第二列至第 $n-1$ 列

⑤ 置 $j = 2$;

⑥ 计算 $\hat{l}_j = \max(\delta, |\gamma_j|^{\frac{1}{2}})$,其中 $\gamma_j = g_{jj} - \sum\limits_{m=1}^{j-1} \bar{l}_{jm}^2$.

⑦ 若 $j = n$,转⑫;

⑧ $\hat{l}_i = \frac{1}{\hat{l}_j}\left(g_{ij} - \sum\limits_{m=1}^{j-1} \bar{l}_{im}\bar{l}_{jm}\right)$　$(i = j+1, \cdots, n)$;

⑨ 计算 $\theta = \max\limits_{j+1 \leq i \leq n} |\hat{l}_i|$;

⑩ 计算 $\bar{l}_{ij}$　$(i = j, \cdots, n)$;

若 $\theta \leq \beta$,令 $\bar{l}_{ij} = \hat{l}_i$　$(i = j, \cdots, n)$;

若 $\theta > \beta$,令

$$\begin{cases} \bar{l}_{jj} = \frac{\theta}{\beta}\hat{l}_j, \\ \bar{l}_{ij} = \frac{\beta}{\theta}\hat{l}_i, \end{cases} \quad (i = j+1, \cdots, n).$$

⑪ 置 $j = j+1$,转⑥

三、计算 $\bar{L}$ 的第 n 列

⑫ 计算 $\bar{l}_{nn} = \max(\delta, |\gamma_n|^{\frac{1}{2}})$,其中 $\gamma_n = g_{nn} - \sum\limits_{m=1}^{n-1} \bar{l}_{nm}^2$.

⑬ 返回主程序.

例 4.2 设 $G=\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}$，$G$ 非正定，试对 G 实行 LL^{T} 分解.

解

$$\xi_1=2,\xi=2,\gamma=1,\beta=1,$$

$$\hat{l}_1=\max(\delta,1)=1,$$

$$\hat{l}_2=\frac{g_{21}}{\hat{l}_1}=2,$$

因为

$$\theta=2,$$

$$\theta>\beta,$$

所以

$$\bar{l}_{11}=2,\quad \bar{l}_{21}=1,$$

$$\bar{l}_{22}=\max\{\delta,|g_{22}-\bar{l}_{21}^2|^{\frac{1}{2}}\}=\max\{\delta,0\}=\delta,$$

故

$$\bar{L}=\begin{bmatrix}2 & 0\\ 1 & \delta\end{bmatrix},$$

$$\bar{G}=\bar{L}\bar{L}^{T}=\begin{bmatrix}4 & 2\\ 2 & 1+\delta^2\end{bmatrix}.$$

例 4.3 设 $G=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$为正定阵，试对它实行 LL^{T} 分解.

解

$$\xi_1=1,\xi=1,\gamma=2,\beta=2,$$

$$\hat{l}_1=\max(\delta,\sqrt{2})=\sqrt{2},$$

$$\hat{l}_2=\frac{g_{21}}{\hat{l}_1}=\frac{1}{\sqrt{2}}=\frac{\sqrt{2}}{2},$$

$$\theta=\max(\hat{l}_1,\hat{l}_2)=\sqrt{2},$$

因为

$$\theta<\beta,$$

所以

$$\bar{l}_{11}=\hat{l}_1=\sqrt{2},\quad \bar{l}_{21}=\frac{\sqrt{2}}{2},$$

$$\bar{l}_{22}=\max\{\delta,|g_{22}-\bar{l}_{21}^2|^{\frac{1}{2}}\}=\max\left\{\delta,\left(2-\frac{1}{2}\right)^{\frac{1}{2}}\right\}=\left(\frac{2}{3}\right)^{\frac{1}{2}},$$

故

$$\bar{L}=\begin{bmatrix}\sqrt{2} & 0\\ \frac{\sqrt{2}}{2} & \left(\frac{3}{2}\right)^{\frac{1}{2}}\end{bmatrix},$$

$$\bar{G}=\bar{L}\bar{L}^{T}=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}=G.$$

算法 4.6（Gill-Murray 算法）

① 送初值 x_0；

② 置 $k=0$;

③ 计算 g_k 和 G_k:

$$g_i^{(k)} = \frac{\partial f(x_k)}{\partial x_i^{(k)}},\ i = 1,2,\cdots,n,$$

$$G_{ij}^{(k)} = \frac{\partial^2 f(x_k)}{\partial x_j^{(k)} \partial x_i^{(k)}},\ i,j = 1,2,\cdots,n;$$

④ 执行算法 4.5,计算 $\bar{L}_k$ 的元素;

⑤ 解方程组

$$\bar{L}_k q_k = -g_k,$$

$$p_k = -G_k^{-1} g_k\text{,或计算 } p_k,$$

⑥ 解方程组

$$\bar{L}_k^{\mathrm{T}} p_k = q_k,$$

⑦ 执行算法 4.2,计算 α_k;

⑧ 令 $x_{k+1} = x_k + \alpha_k p_k$;

⑨ 执行第 2 章算法 2.3,不满足 H - 终止法则转⑩,否则转⑪;

⑩ 置 $k=k+1$,转③;

⑪ Stop.

第5章　共轭方向法

前面叙述的优化算法，如梯度法(最速下降法)，Newton 法都有自身的局限性. 梯度法程序简单，计算量小，但收敛速度很慢(线性敛速)；Newton 法虽然收敛速度较快(二阶敛速)但对初值的依赖较大，且有第 4 章中所陈述的种种缺陷.

共轭方向法似乎介于梯度法和 Newton 法之间，它具有超线性敛速，而且迭代公式比较简单，只用到目标函数及其梯度值，避免了二阶导数的计算，从而减少了计算量和存储量，因此它是比较有效的最优化方法.

5.1　共轭方向

定义 5.1　设 $\boldsymbol{A}=(a_{ij})_n$ 为对称阵，$\boldsymbol{p}_i$ $(i=0,1,\cdots,m,m<n)$ 为 n 维向量组，若

$$(\boldsymbol{p}_i,\boldsymbol{p}_j)_A=\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_j=0\quad(i\neq j),\tag{5-1}$$

则称 $\boldsymbol{p}_i$ $(i=0,1,\cdots,m)$ 为 $\boldsymbol{A}$-共轭向量系.

显然，若 $\boldsymbol{A}=\boldsymbol{I}$，则变成正交的定义，可见共轭是正交的推广，或者叫做 $\boldsymbol{A}$-正交.

下面叙述 $\boldsymbol{A}$-共轭向量系的一些性质.

定理 5.1　设(1) $\boldsymbol{A}=(a_{ij})_n$ 为对称正定阵；

(2) $\boldsymbol{p}_i\neq 0$ $(i=0,1,\cdots,m)$ 为 $\boldsymbol{A}$-共轭向量系，

则向量系 $\boldsymbol{p}_i$ $(i=0,1,\cdots,m)$ 是线性独立的.

证　设相反，$\{\boldsymbol{p}_i\}$ 线性相关，则存在不全为 0 的 c_i $(i=0,1,\cdots,m)$ 使得

$$\sum_{i=0}^{m}c_i\boldsymbol{p}_i=0,\tag{5-2}$$

以 $\boldsymbol{A}$ 分别左乘(5-2)式得

$$\sum_{i=0}^{m}c_i\boldsymbol{A}\boldsymbol{p}_i=0,\tag{5-3}$$

从而有

$$\boldsymbol{p}_j^{\mathrm{T}}\left(\sum_{i=0}^{m}c_i\boldsymbol{A}\boldsymbol{p}_i\right)=0,\quad j=0,1,\cdots,m,$$

即

$$\sum_{i=0}^{m}c_i\boldsymbol{p}_j^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_i=0,\quad j=0,1,\cdots,m,$$

由假设(2)有

$$c_j p_j^{\mathrm{T}} A p_j = 0, \quad j = 0,1,\cdots,m,$$

但是　$p_j \neq 0,\quad p_j^{\mathrm{T}} A p_j > 0$($A$ 正定).

故　$c_j = 0,\quad j = 0,1,\cdots,m,$

这与假设矛盾,得证.

定理 5.2　设(1) $A = (a_{ij})_n$;

(2) $p_i \neq 0 \quad (i = 0,1,\cdots,n-1)$为 A - 共轭;

(3) v 与 $p_i \quad (i = 0,1,\cdots,n-1)$为 A - 共轭,

则 $v = 0$.

证　由定理 5.1,$p_i \quad (i = 0,1,\cdots,n-1)$线性独立,并构成 $\mathbf{R}^n$ 上的基底,故若 $v \neq 0$,则

$$v = \sum_{j=0}^{n-1} c_j p_j, \tag{5-4}$$

其中 $c_j \quad (j = 0,1,\cdots,m)$不全为 0,用 $p_i^{\mathrm{T}} A \quad (i = 0,1,\cdots,n-1)$左乘(5-4)并由假设(2)得

$$\begin{aligned} p_i^{\mathrm{T}} A v &= \sum_{j=0}^{n-1} c_j p_i^{\mathrm{T}} A p_j \\ &= c_i p_i^{\mathrm{T}} A p_i \quad (i = 0,1,\cdots,n-1), \end{aligned} \tag{5-5}$$

由假设(3)知　$p_i^{\mathrm{T}} A v = 0.$

据(5-5)　$c_i p_i^{\mathrm{T}} A p_i = 0 \quad (i = 0,1,\cdots,n-1),$

由(5-4)知,若 $v \neq 0$,则 $c_i \quad (i = 0,1,\cdots,n-1)$不全为 0,

故有　$p_i^{\mathrm{T}} A p_i = 0 \quad (i = 0,1,\cdots,n-1),$

这与假设矛盾. 所以　$v = 0.$

定理 5.2 指出,若 A 对称正定,则 $\mathbf{R}^n$ 中至多有 n 个不同的 A - 共轭方向.

定理 5.3　设(1) A 对称正定;

(2) $p_i \neq 0 \quad (i = 0,1,\cdots,n-1)$为 A - 共轭方向系;

(3) $v \in \mathbf{R}^n$ 为任意向量,

则

$$v = \sum_{i=0}^{n-1} \frac{p_i^{\mathrm{T}} A v}{p_i^{\mathrm{T}} A p_i} p_i. \tag{5-6}$$

证　由定理 5.1,向量系 $p_i \neq 0 \quad (i = 0,1,\cdots,n-1)$,于是构成 $\mathbf{R}^n$ 中的一个基底. 故对 $\forall v \in \mathbf{R}^n$ 有

$$v = \sum_{j=0}^{n-1} c_j p_j, \tag{5-7}$$

用 $p_i^T A$ $(i=0,1,\cdots,n-1)$ 左乘(5-7)得

$$p_i^T A v = \sum_{j=0}^{n-1} c_j p_i^T A p_j = c_i p_i^T A p_i,$$

故
$$c_i = \frac{p_i^T A v}{p_i^T A p_i},$$

将 c_i 代入(5-7)式即得(5-6).

5.2 共轭方向法

由于非线性目标函数在极小点附近可以用正定二次函数去逼近,因此二次函数的极小化的有效算法是构造一般非线性目标函数极小化算法的基础. 为此,我们首先讨论正定二次函数的共轭方向法,然后推广到一般的非线性函数上去.

设二次正定函数

$$f(x) = \frac{1}{2}x^T A x + b^T x + c,$$

其中 A 对称正定,$g(x) = \nabla f(x) = Ax + b$,显然,$f(x)$ 在 $\mathbf{R}^n$ 中唯一的全局极小点为 $x^* = -A^{-1}b$,且

$$g(x^*) = 0.$$

在第 2 章所述的下降算法 2.1 中,若令下降方向 p_k $(k=0,1,\cdots)$ 为共轭方向,并且 α_k 由精确的一维寻查确定,便得到关于正定二次函数的共轭方向法.

算法 5.1 给出 x^* 的初始近似 x_0.

① 计算 $g_0 = g(x_0)$;

② 构造 p_0 满足 $g_0^T p_0 < 0$;

③ 令 $k=0$;

④ 计算 α_k,满足

$$f(x_k + \alpha_k p_k) = \min_{\alpha > 0} f(x_k + \alpha p_k);$$

⑤ 令 $x_{k+1} = x_k + \alpha_k p_k$;

⑥ 检查 $k = n-1$?

是,输出 x_{k+1},停止计算;

⑦ 计算 p_{k+1} 满足 $p_{k+1}^T A p_j = 0$ $(j=0,1,\cdots,k)$;

⑧ 令 $k=k+1$,转④.

算法 5.1 具有如下两个特点:

(1) 一维寻查的步长因子 α_k 可用公式表出:

$$\alpha_k = -\frac{p_k^T g_k}{p_k^T A p_k}, \tag{5-8}$$

事实上，将二次函数 $f(x)$ 在 x_k 处展开，有

$$f(x_k + \alpha p_k) = f(x_k) + \alpha p_k^T g_k + \frac{1}{2}\alpha^2 p_k^T A p_k = \phi(\alpha).$$

由 $\phi'(\alpha_k)=0$，有

$$p_k^T g_k + \alpha_k p_k^T A p_k = 0, \tag{5-9}$$

由(5-9)解出 α_k 即得(5-8)式. 在(5-8)式中，若令 $p_k = -g_k$ 即得第二章(2-17)式.

(2) 算法具有二次收敛性，即对任意选取的初值 x_0，至多迭代 n 步可达 $f(x)$ 的全局极小点 x^*. 详述于下面的定理.

定理 5.4(共轭方向法具二次收敛性)　设

(1) 任取 $x_0 \in \mathbf{R}^n$;

(2) p_k　$(k=0,1,\cdots,n-1)$ 为 A-共轭向量系;

(3) x_k　$(k=1,2,\cdots)$ 由算法 5.1 产生，则存在某个 $m \leq n$，使 $x_m = x^*$.

证　如前所述

$$\alpha_k = -\frac{p_k^T g_k}{p_k^T A p_k} = -\frac{p_k^T (A x_k + b)}{p_k^T A p_k}. \tag{5-10}$$

而在(5-10)中，

$$p_k^T(Ax_k + b) = p_k^T A x_k + p_k^T b = p_k^T A(x_0 + \sum_{j=0}^{k-1}\alpha_i p_j) + p_k^T b$$

$$= p_k^T A x_0 + \sum_{j=0}^{k-1}\alpha_j p_k^T A p_j + p_k^T b = p_k^T A x_0 + p_k^T b,$$

所以

$$\alpha_k = -\frac{p_k^T A x_0}{p_k^T A p_k} - \frac{p_k^T b}{p_k^T A p_k}. \tag{5-11}$$

于是由(5-11)得

$$x_n = x_0 + \sum_{j=0}^{n-1}\alpha_j p_j = x_0 - \sum_{j=0}^{n-1}\frac{p_j^T A x_0}{p_j A p_j}p_j - \sum_{j=0}^{n-1}\frac{p_j^T A(A^{-1}b)}{p_j^T A p_j}p_j, \tag{5-12}$$

据定理 5.3 简化(5-12)式为

$$x_n = x_0 - x_0 - A^{-1}b = -A^{-1}b = x^*.$$

如果对于给定的 x_0 及 p_0，恰好

$$\alpha_k = 0 \quad (k = m+1,\cdots,n-1),$$

则

$$x_m = x^*.$$

上述定理表明，对于任意选取的 x_0，至多迭代 n 次便达到 $f(x)$ 的极

小点，在特殊情况下，算法还可提前终止.

综上所述，对于正定二次函数的共轭方向法（算法 5.1），可示意如下：

给定 x_0，选取下降方向 p_0 （$g_0^T p_0 < 0$），计算

$$x_1 = x_0 + \alpha_0 p_0,$$

其中

$$\alpha_0 = -\frac{g_0^T p_0}{p_0^T A p_0}.$$

从 x_1 出发，选取共轭方向 p_1（$p_1^T A p_0 = 0$），计算

$$x_2 = x_1 + \alpha_1 p_1,$$

其中

$$\alpha_1 = -\frac{g_1^T p_1}{p_1^T A p_1}.$$

一般地，从 x_k 出发，选取共轭方向 p_k （$p_k^T A p_j = 0, j = 0,1,\cdots,k-1$），计算

$$x_{k+1} = x_k + \alpha_k p_k \quad (k = 0,1,\cdots,n-1),$$

其中

$$\alpha_k = -\frac{g_k^T p_k}{p_k^T A p_k}.$$

将算法 5.1 中的终止法则作适当的修改便可得到适合于一般函数的共轭方向法.

下面进一步讨论共轭方向法的性质，这些性质中，有的为构造具体的共轭方向提供工具，有的则为共轭方向法奠定理论基础.

定理 5.5 设(1) $x_0 \in \mathbf{R}^n$ 任给；

(2) $f(x)$ 为正定二次函数；

(3) p_i （$i = 0,1,\cdots,k$）为 A - 共轭系；

(4) x_i （$i = 0,1,\cdots,k+1$）由算法 5.1 产生，

则 $g_{k+1}^T p_i = 0 \quad (i = 0,1,\cdots,k)$，

即 $g_{k+1} \perp L_{k+1} = \mathrm{Span}\{p_0, p_1, \cdots, p_k\}$.

需要指出，只要 α_k 是由一维寻查确定的，则由第 3 章(3 - 2)式，$g_{k+1}^T p_k = 0$ （$k \geqslant 0$）不论 $f(x)$ 是否二次函数，也不论 $\{p_k\}$ 是否 A 共轭系均成立. 但若 $f(x)$ 为正定二次函数，且 $\{p_k\}$ 为 A 共轭系，则有更强的结果，即定理 5.5 的结论（二维情形如图 5 - 1 所示）.

证 因为 $g(x) = Ax + b$.

所以

$$g_{k+1}^T p_i = (A x_{k+1} + b)^T p_i = \left(A\left(x_{i+1} + \sum_{j=i+1}^{k} \alpha_j p_j\right) + b\right)^T p_i,$$

$$= g_{i+1}^T p_i + \sum_{j=i+1}^{k} \alpha_j p_j^T A p_i \quad (0 \leqslant i < k), \tag{5 - 13}$$

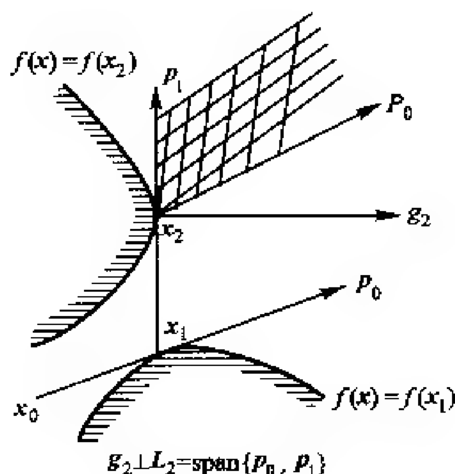


图 5-1

又因
$$g_{i+1}^{\mathrm{T}} p_i = 0 \qquad (i \geqslant 0),$$
且由假设(3)
$$p_j A p_i = 0 \qquad (i \neq j),$$
于是由(5-13),得
$$g_{k+1}^{\mathrm{T}} p_i = 0 \qquad (i = 0, \cdots, k-1), \tag{5-14}$$
而
$$g_{k+1}^{\mathrm{T}} p_k = 0, \tag{5-15}$$
综合(5-14)和(5-15),得
$$g_{k+1}^{\mathrm{T}} p_i = 0 \qquad (i = 0, \cdots, k), \tag{5-16}$$

定理 5.6 设(1) $f:\mathbf{R}^n \to \mathbf{R}^1$ 在 $\mathbf{R}^n$ 是严格凸函数;

(2) $f(x)$在 $\mathbf{R}^n$ 内有一阶连续偏导数;

(3) x^* 为$f(x)$的无约束极小点;

(4) $x_0 \in \mathbf{R}^n$ 任给;

(5) $p_i \ (i=0,1,\cdots,k-1)$在 $\mathbf{R}^n$ 内是线性独立的,并记
$$x_k = x_0 + \sum_{i=0}^{k-1} \alpha_i p_i,$$
$$M_k = \left\{ x \in \mathbf{R}^n \mid x = x_0 + \sum_{i=0}^{k-1} \alpha_i p_i \right\},$$
则 $f(x_k) = \min\limits_{x \in M_k} f(x)$的充要条件是
$$g_k \perp L_k = \mathrm{span}\{p_0, p_1, \cdots, p_{k-1}\}. \tag{5-17}$$

二维情形下(如图 5-2 所示),定理的结论可表为:$f(x_2) = \min\limits_{x \in M_2} f(x)$的充要条件是
$$g(x_2) \perp M_2 \qquad (M_2 \mathbin{/\!/} L_2).$$

证 必要性 设 x_k 是$f(x)$在线性流形 $x = x_0 + z$ 上的极小点,其中

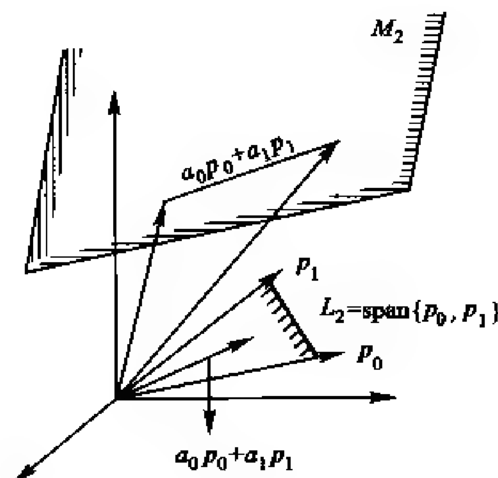


图 5－2

$z \in L_k = \text{span}\{p_0, p_1, \cdots, p_{k-1}\}$，则 $f(x)$ 在 x_k 点沿任何方向导数为 0，

即
$$g(x_k)^{\mathrm{T}} z = 0, \tag{5-18}$$

由此即得(5－17).

充分性　若 $x_k \in M_k$ 满足

$$g_k^{\mathrm{T}} p_i = 0 \quad (i = 0, 1, \cdots, k-1),$$

现欲证明 x_k 是 $f(x)$ 在 M_k 上的严格极小点，即对于

$$\forall x \in M_k \quad (x \neq x_k),$$

有
$$f(x_k) < f(x).$$

设相反，对 $\forall x \in M_k \quad (x \neq x_k)$ 有

$$f(x) \leqslant f(x_k),$$

因为
$$x \neq x_k,$$

由 M_k 的定义有

$$\begin{cases} x = x_0 + \sum_{i=0}^{k-1} \alpha_i p_i; \\ x_k = x_0 + \sum_{i=0}^{k-1} \alpha'_i p_i, \end{cases}$$

故
$$x - x_k = \sum_{i=0}^{k-1} \alpha''_i p_i,$$

即
$$x = x_k + \sum_{i=0}^{k-1} \alpha''_i p_i \quad (\alpha''_i = \alpha_i - \alpha'_i).$$

再由假设(1)、(2)及第 1 章定理 1.11，$f(x)$ 严格凸的充要条件为

$$f(x) > f(x_k) + g_k^{\mathrm{T}}(x - x_k) = f(x_k) + g_k^{\mathrm{T}} \sum_{i=0}^{k-1} \alpha''_i p_i = f(x_k),$$

这与假设矛盾,定理得证.

由定理5.5及定理5.6立即可得二次函数共轭方向法的一个极为重要的性质.

定理5.7(扩展子空间定理) 对于正定二次函数$f(x)$,由算法5.1所产生的点x_k是$f(x)$在$M_k = \{x \in \mathbf{R}^n \mid x = x_0 + z, z \in \text{span}(p_0, p_1, \cdots, p_{k-1})\}$上的唯一极小点.

事实上,本定理既符合定理5.5的条件,又符合定理5.6的条件.由定理5.5的(5-16),有

$$g_k \perp L_k = \text{span}\{p_0, p_1, \cdots, p_{k-1}\}.$$

又由定理5.6的(5-17),得

$$f(x_k) = \min_{x \in M_k} f(x).$$

注 (1) 定理5.7表明,对于正定二次函数$f(x)$,用共轭方向法求极小点$x^* \in \mathbf{R}^n$的过程实际上是子空间$\text{span}\{p_0, p_1, \cdots, p_{k-1}\}$的逐次扩展过程:

x_1是$f(x)$在M_1上的极小点(M_1是过x_0且平行于p_0的直线);

x_2是$f(x)$在M_2上的极小点(M_2是过x_0且平行于子空间$\text{span}\{p_0, p_1\}$的超平面)(见图5-2);

$x_n = x^*$是$f(x)$在M_n上的极小点(M_n是过x_0且平行于子空间$\text{span}\{p_0, p_1, \cdots, p_{n-1}\}$的超平面);

(2) 若$x_0 = 0$,则x_k是$f(x)$在$L_k = \text{span}\{p_0, p_1, \cdots, p_{k-1}\}$上的极小点.而$x_n = x^*$是$f(x)$在$\mathbf{R}^n$上的极小点.

若$x^* \in \text{span}\{p_0, p_1, \cdots, p_{m-1}\} = L_m$,则$x_m = x^*$ $(m < n)$,这就是定理5.4中所提到的算法提前终止的情形.

5.3 共轭梯度法

5.3.1 正定二次函数的基本算法

1. 共轭方向的构造

迄今为止,我们只介绍了共轭方向的性质,并未给出构造共轭方向的具体方法.下面我们仿照Gran-schmidt正交化过程,介绍一种构造共轭方向的方法:

设 $$f(x) = \frac{1}{2}x^{\mathrm{T}}Ax + b^{\mathrm{T}}x + c \quad (A\text{ 正定}),$$

给定x_0,取$p_0 = -g_0$;

计算

$$\begin{cases} x_1 = x_0 + \alpha_0 p_0, & \text{其中 } \alpha_0 \text{ 由下式确定} \\ \alpha_0 = -\dfrac{g_0^T p_0}{p_0^T A p_0}. \end{cases}$$

对于 x_1，取 $p_1 = -g_1 + \beta_0^{(1)} p_0$，其中 $\beta_0^{(1)}$ 利用 $(p_1, p_0)_A = 0$ 确定如下：

$$\beta_0^{(1)} = \frac{(g_1, p_0)_A}{(p_0, p_0)_A};$$

计算

$$\begin{cases} x_2 = x_1 + \alpha_1 p_1, & \text{其中 } \alpha_1 \text{ 由下式确定} \\ \alpha_1 = -\dfrac{g_1^T p_1}{p_1^T A p_1}. \end{cases}$$

对于 x_2，取 $p_2 = -g_2 + \beta_0^{(2)} p_0 + \beta_1^{(2)} p_1$，其中 $\beta_0^{(2)}, \beta_1^{(2)}$ 利用 $(p_2, p_j)_A = 0, j = 0,1$ 确定如下：

$$\beta_j^{(2)} = \frac{(g_2, p_j)_A}{(p_j, p_j)_A}, \quad j = 1,2;$$

计算

$$\begin{cases} x_3 = x_2 + \alpha_2 p_2, & \text{其中 } \alpha_2 \text{ 由下式确定} \\ \alpha_2 = -\dfrac{g_2^T p_2}{p_2^T A p_2}. \end{cases}$$

....

若已构造出 A 共轭系

$$p_0, p_1, \cdots, p_{k-1} \text{ 及 } x_k,$$

对于 x_k，取 $p_k = -g_k + \sum_{j=0}^{k-1} \beta_j^{(k)} p_j$， (5 - 19)

利用 $(p_k, p_j)_A = 0 \quad (j = 0,1,\cdots,k-1)$，确定

$$\beta_j^{(k)} = \frac{(g_k, p_j)_A}{(p_j, p_j)_A} \quad (j = 0,1,\cdots,k-1); \tag{5 - 20}$$

计算

$$\begin{cases} x_{k+1} = x_k + \alpha_k p_k, & \text{其中 } \alpha_k \text{ 由下式确定} \\ \alpha_k = -\dfrac{g_k^T p_k}{p_k^T A p_k}. \end{cases}$$

重复上述计算，直至 $k = n - 1$ 为止：

2. p_k 计算公式的简化

应当注意，在(5 - 19)中，可以证明

$$\beta_j^{(k)} = 0, \quad j = 0,1,\cdots,k-2,$$

从而可得 $$p_k = -g_k + \beta_{k-1}^{(k)} p_{k-1}.$$

事实上，利用 $g(x) = Ax + b$，可得

$$g_{j+1} - g_j = (Ax_{j+1} + b) - (Ax_j + b) = A(x_{j+1} - x_j) = A\alpha_j p_j$$

$(j = 0, 1, \cdots, k)$，

即 $$Ap_j = \frac{1}{\alpha_j}(g_{j+1} - g_j), \tag{5-21}$$

利用(5－21)改写 $\beta_j^{(k)}$ 的表达式(5－20)，其分子为

$$(g_k, p_j)_A = (g_k, Ap_j) = \frac{1}{\alpha_j}(g_k, g_{j+1}) - \frac{1}{\alpha_j}(g_k, g_j)$$

$$(j = 0, 1, \cdots, k-1), \tag{5-22}$$

又由(5－19) $$g_j = -p_j + \sum_{i=0}^{j-1} \beta_i^{(j)} p_i,$$

于是 $$(g_k, g_j) = -(g_k, p_j) + \sum_{i=0}^{j-1} \beta_i^{(j)} (g_k, p_i). \tag{5-23}$$

当 $j \leqslant k-1$ 时，$(g_k, p_j) = 0$，并且

$$\sum_{i=0}^{j-1} \beta_i^{(j)} (g_k, p_i) = 0,$$

由(5－23)知 $(g_k, g_j) = 0$，$j \leqslant k-1$.

又由 $(g_k, g_{j+1}) = (g_k, g_j) = 0 \quad (j \leqslant k-2)$ 及(5－22).

得 $$(g_k, p_j)_A = 0, \quad j \leqslant k-2,$$

据(5－19)和(5－20)，分别有

$$\beta_j^{(k)} = 0, \quad j = 0, 1, \cdots, k-2,$$

$$p_k = -g_k + \beta_{k-1}^{(k)} p_{k-1},$$

或 $$p_k = -g_k + \beta_{k-1} p_{k-1},$$

其中 $$\beta_{k-1}^{(k)} \xlongequal{\text{记}} \beta_{k-1} = \frac{g_k^{\mathrm{T}} A p_{k-1}}{p_{k-1}^{\mathrm{T}} A p_{k-1}}.$$

3．共轭梯度法

根据 1、2 两款的讨论，归纳共轭梯度法如下：

算法 5.2

① 任给 $x_0 \in \mathbf{R}^n$；

② 取 $p_0 = -g_0 = b - Ax_0$；

③ 对 $k = 0, 1, \cdots, n-1$，分别依次计算

$$\alpha_k = -\frac{g_k^{\mathrm{T}} p_k}{p_k^{\mathrm{T}} A p_k},$$

$$x_{k+1} = x_k + \alpha_k p_k,$$

$$g_{k+1} = g_k + \alpha_k A p_k,$$

$$\beta_k = \frac{g_{k+1}^T p_k}{p_k^T A p_k},$$

$$p_{k+1} = -g_{k+1} + \beta_k p_k.$$

由 5.3.1 第 1 款知，我们是在假定 $\{p_k\}$ 为共轭方向的前提下来构造 $\{p_k\}$ 的，因此由算法 5.2 所生成的方向系 $\{p_k\}$ 为共轭方向系，从而经 n 步迭代所得 x_n 为正定二次函数 $f(x)$ 的全局极小点.

5.3.2 基本性质

定理 5.8　算法 5.2 中所出现的两组向量系 $\{p_k\}$，$\{g_k\}$ 有如下关系：

(1) $\{p_k\}$ 是 A 共轭系；

(2) $\{g_k\}$ 是直交系；

(3) $\{p_k\}$，$\{g_k\}$，满足关系式

$$(g_k, p_j) = 0, \quad j < k,$$

$$(p_k, g_j) = (p_k, g_k), \quad j \leqslant k;$$

(4) $p_j(j=0,1,\cdots,m)$ 是 x_k 处的下降方向.

证　(1) 如前所述，结论(1)显然成立；

(2) 在简化 p_k 的计算公式的推导中，有

$$(g_k, g_j) = 0, \quad (j \leqslant k-1 \text{ 或 } j < k),$$

故结论(2)亦成立；

(3) $g_k^T p_j = 0 \quad (j<k)$ 已在定理 5.6 中得证，现待证

$$p_k^T g_j = p_k^T g_k \quad (j \leqslant k).$$

事实上，　$p_k^T g_{j+1} = p_k^T(g_j + \alpha_j A p_j) = (p_k^T g_j + \alpha_j p_k^T A p_j).$

若 $j<k$，据(1)之结果有 $p_k^T A p_j = 0$，故有

$$p_k^T g_{j+1} = p_k^T g_j,$$

令 $j=1,2,\cdots,k-1$，有

$$p_k^T g_1 = p_k^T g_2 = \cdots = p_k^T g_{k-1} = p_k^T g_k,$$

于是　$$p_k^T g_j = p_k^T g_k \quad (j \leqslant k).$$

(4) $$g_k^T p_j = g_k^T(-g_j + \beta_{j-1} p_{j-1}) \quad (j \leqslant k)$$

$$= -g_j^T g_j + \beta_{j-1} g_j^T p_{j-1} = -\|g_j\|_2^2 < 0 \quad (j=0,1,\cdots,m).$$

由第一章定义 1.11 知 $p_j \quad (j=0,1,\cdots,m)$ 是 x_k 处的下降方向.

定理 5.9　在算法 5.1 中，关于系数

$$\alpha_k = -\frac{g_k^T p_k}{p_k^T A p_k},$$

$$\beta_k = \frac{g_{k+1}^T A p_k}{p_k^T A p_k},$$

还可以用如下交替公式

$$\alpha_k = \frac{\|g_k\|_2^2}{p_k^T A p_k},$$

$$\beta_k = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|g_k\|_2^2} \quad \text{(Polak - Ribiere 公式)}$$

$$= \frac{\|g_{k+1}\|_2^2}{\|g_k\|_2^2} \quad \text{(Fletcher - Reeres 公式)}.$$

证　因为　$-g_k^T p_k = -g_k^T(-g_k - \beta_{k-1} p_{k-1}) = g_k^T g_k = \|g_k\|_2^2,$

所以
$$\alpha_k = \frac{\|g_k\|_2^2}{p_k^T A p_k}$$

又因为　$g_{k+1}^T A p_k = \frac{1}{\alpha_k} g_{k+1}^T (g_{k+1} - g_k) = \frac{1}{\alpha_k} g_{k+1}^T (g_{k+1} - g_k) = \frac{1}{\alpha_k} \|g_{k+1}\|_2^2,$

$$p_k^T A p_k = \frac{1}{\alpha_k} p_k^T (g_{k+1} - g_k) = \frac{1}{\alpha_k} g_k^T g_k = \frac{1}{\alpha_k} \|g_k\|_2^2,$$

所以
$$\beta_k = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|g_k\|_2^2} = \frac{\|g_{k+1}\|_2^2}{\|g_k\|_2^2}$$

根据上述性质，改进算法 5.2 如下：

算法 5.3

① 任给 $x_0 \in \mathbf{R}^n$；

② 取 $p_0 = -g_0 = b - Ax_0$；

③ 对 $k = 0, 1, \cdots, n-1$，计算

$$\alpha_k = -\frac{g_k^T p_k}{p_k^T A p_k} = \frac{\|g_k\|_2^2}{p_k^T A p_k}$$

$$x_{k+1} = x_k + \alpha_k p_k,$$

$$g_{k+1} = g_k + \alpha_k A p_k,$$

$$\beta_k = \frac{g_{k+1}^T p_k}{p_k^T A p_k} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|g_k\|_2^2} = \frac{\|g_{k+1}\|_2^2}{\|g_k\|_2^2},$$

$$p_{k+1} = -g_{k+1} + \beta_k p_k.$$

5.3.3　一般函数的共轭梯度法

1. PRP 算法(Polyok-Ribiere-Polak)

算法 5.4

① 取初始点 x_0；

② 置取 $p_0=-g_0, k=0$;

③ 求 α_k,使 $f(x_k+\alpha_k p_k)=\min f(x_k+\alpha p_k)$;

④ 计算 $x_{k+1}=x_k+\alpha_k p_k$;

⑤ 计算 $g_{k+1}=\nabla f(x_{k+1})$;

⑥ 计算 $\beta_k=\dfrac{g_{k+1}^{T}(g_{k+1}-g_k)}{\|g_k\|_2^2}$;

⑦ 计算 $p_{k+1}=-g_{k+1}+\beta_k p_k$;

⑧ 置 $k=k+1$,转③.

显然,对于一般的连续可微函数来说,PRP 算法构造的寻查方向 p_k 应满足

$$(g_k,p_k)=(g_k,-g_k+\beta_{k-1}p_{k-1})=(g_k,-g_k)=-\|g_k\|_2^2.$$

这表明当 $g_k\neq 0$,则在 x_k 处沿 p_k 的方向导数取负值,从而 p_k 总是下降方向.

对于正定二次函数来说,由于

$$(g_{k+1},p_k)_A=g_{k+1}^{T}(g_{k+1}-g_k)=\|g_{k+1}\|_2^2,$$

所以

$$\beta_k=\frac{\|g_{k+1}\|_2^2}{\|g_k\|_2^2},$$

而在算法 PRP 中,若 β_k 的计算公式用上式代替,这就是所谓的 FR 算法(Fietcher - Reves),这两个算法对于正定二次函数来说是等价的,对于一般目标函数,两个算法每进行一次迭代的计算量也大体相同,但计算效果有时却会有不小的差异. 一般来说,PRP 算法优于 FR 算法,有兴趣的读者可以参阅参考文献[24].

还需指出,若 $f(x)$ 满足一定条件,则 PRP 算法具有全局收敛性,即从任意初始点 x_0 出发,按算法产生的点列,或者终止或者收敛于 $f(x)$ 在 $\mathbf{R}^n$ 上的唯一极小点 x^*,而且还可进一步证明其收敛速度为线性收敛.

2. PRP 算法的返转策略

对正定二次函数来说,由于 PRP 算法是一种共轭方向法,因此具有二次收敛性,即从任意初值出发,迭代 n 步后可达到极小点,而实际情况是,算法可能提前终止,即迭代 m 步($m<n$)后,x_m 便是 $f(x)$ 的极小点,此时 $x^*\in \operatorname{span}\{p_0,p_1,\cdots,p_{m-1}\}=L_m$,或者由于舍入误差的影响,迭代也可能退后. 此时需要重新调整方向.

至于一般函数,如前所述,PRP 算法很可能只是线性收敛,如果在迭代过程中,目标函数始终不能用二次函数较好地近似,那么收敛慢的原因可归结为目标函数"性态不好",但已经有例于表明,即使对于性态很好的函数,算法仍可能收敛很慢,这种情况的出现,只能归结于我们的策略

不佳，即当迭代已进入二次函数的区域后，未能及时找到几个共轭方向，此时亦需重新调整初始寻查方向.

我们知道，共轭梯度法构造共轭方向系时，关键步骤之一是取负梯度方向作为第一个寻查方向，因此当算法出现收敛缓慢时，可采取重新置负梯度方向的措施来加速收敛，这就是所谓的返转策略，这个过程示意如下：

					$k=n-1$
			p_0	p_1	$\cdots$	p_{n-1}
		x_0	x_1	x_2	$\cdots$	x_n
	$p_0^{(1)}=-g_n$			$p_1^{(1)}$	$\cdots$	$p_{n-1}^{(1)}$
$x_0^{(1)}=x_n$	$x_1^{(1)}$			$x_2^{(1)}$	$\cdots$	$x_n^{(1)}$
$\cdots$				$\cdots$		$\cdots$

实践表明，进行 $n+1$ 次迭代后再返转比进行 n 次迭代后再返转更为有利，故可改为如下过程：

						$k=n-1$	$k=n$
			p_0	p_1	$\cdots$	p_{n-1}	p_n
		x_0	x_1	x_2	$\cdots$	x_n	x_{n+1}
	$p_0^{(1)}=-g_{n+1}$			$p_1^{(1)}$	$\cdots$	$p_{n-1}^{(1)}$	$p_n^{(1)}$
$x_0^{(1)}=x_{n+1}$	$x_1^{(1)}$			$x_2^{(1)}$	$\cdots$	$x_n^{(1)}$	$x_{n+1}^{(1)}$
	$\cdots$				$\cdots$	$\cdots$	

算法 5.5（n 步重新开始的 PRP 算法）

① 取 x_0 置 $k=0$；

② 计算 $g_k=\nabla f(x_k)$；

③ 若 $g_k=0$，停止计算，否则置

$$p_k=-g_k+\beta_{k-1}p_{k-1},$$

其中

$$\beta_{k-1}=\begin{cases}0, & \text{当 } k \text{ 是 } n \text{ 的整数倍时}\\ \dfrac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{\|g_{k-1}\|_2^2}; \end{cases}$$

④ 求 α_k，使 $f(x_k+\alpha_k p_k)=\min\limits_{\alpha\geqslant 0} f(x_k+\alpha p_k)$；

⑤ 置 $x_{k+1}=x_k+\alpha_k p_k$；

⑥ 置 $k=k+1$，转②.

框图见图 5－3.

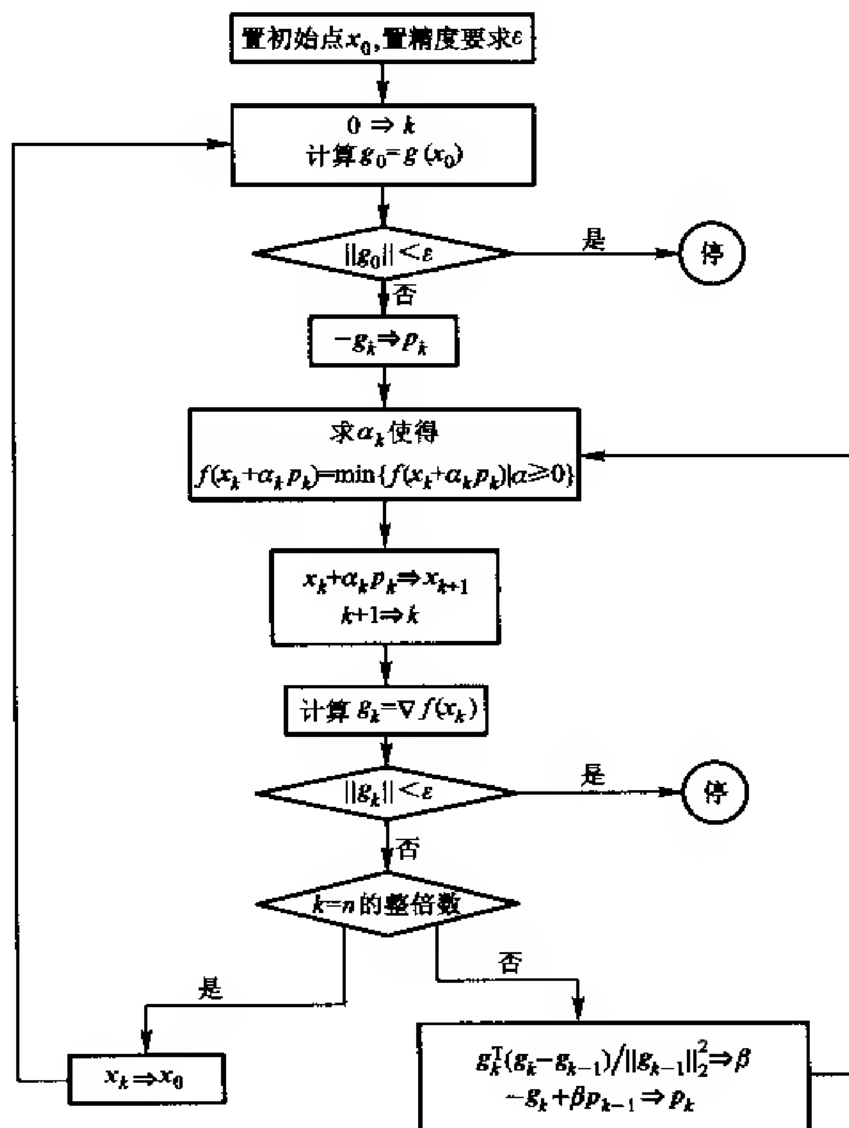


图 5－3

第 6 章　拟 Newton 法

6.1　尺度矩阵意义下的最速下降方法

本章介绍一种在实际应用中更具有可操作性的改进 Newton 法，即拟 Newton 法，或称变尺度方法. 首先我们扩展下降方向(第 1 章定义 1.11)的概念，引入尺度矩阵意义下的下降方向及最速下降方向.

设 $f:S\subset \mathbf{R}^n\to\mathbf{R}^1$，$f(\boldsymbol{x})$ 在点 $\hat{\boldsymbol{x}}\in S$ 处可微，$\boldsymbol{p}\in\mathbf{R}^n$ 是已知的非零向量，$\boldsymbol{A}$ 为正定对称矩阵，对 $\forall\varepsilon>0$ 有

$$\lim_{\varepsilon\to 0}\frac{f(\hat{\boldsymbol{x}}+\varepsilon\boldsymbol{p})-f(\hat{\boldsymbol{x}})}{\varepsilon}=<\nabla f(\hat{\boldsymbol{x}}),\boldsymbol{p}>_{\boldsymbol{A}}=\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}.$$

据 Cauchy-Schwarz 不等式有

$$\langle\nabla f(\hat{\boldsymbol{x}}),\boldsymbol{p}\rangle_{\boldsymbol{A}}^2\leqslant\|\nabla f(\hat{\boldsymbol{x}})\|_{\boldsymbol{A}}^2\cdot\|\boldsymbol{p}\|_{\boldsymbol{A}}^2,$$

即

$$|\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}|\leqslant\|\nabla f(\hat{\boldsymbol{x}})\|_{\boldsymbol{A}}\cdot\|\boldsymbol{p}\|_{\boldsymbol{A}}.\qquad(6-1)$$

当取 $\boldsymbol{p}=-\boldsymbol{A}^{-1}\nabla f(\hat{\boldsymbol{x}})$ 时，因为 $<\nabla f(\hat{\boldsymbol{x}}),\boldsymbol{p}>_{\boldsymbol{A}}=\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}=-\|\nabla f(\hat{\boldsymbol{x}})\|_2^2<0$，据下降方向的定义，称 $\boldsymbol{p}$ 为 $f(\boldsymbol{x})$ 在 $\hat{\boldsymbol{x}}\in S$ 处(在 $\boldsymbol{A}$ 尺度矩阵意义下)的下降方向；另一方面当 $\boldsymbol{p}=-\boldsymbol{A}^{-1}\nabla f(\hat{\boldsymbol{x}})$ 时，不等式(6－1)成立等号，$|\nabla f(\hat{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}|$ 达最大值，因此我们称 $\boldsymbol{p}$ 为 $f(\boldsymbol{x})$ 在 $\hat{\boldsymbol{x}}\in S$ 处(在 $\boldsymbol{A}$ 尺度矩阵意义下)的最速下降方向. 显然，当 $\boldsymbol{A}=\boldsymbol{E}$(单位矩阵)时，$\boldsymbol{p}=-\boldsymbol{E}^{-1}\nabla f(\hat{\boldsymbol{x}})=-\nabla f(\hat{\boldsymbol{x}})$ 是 $f(\boldsymbol{x})$ 在 $\hat{\boldsymbol{x}}\in S$ 处的最速下降方向($\boldsymbol{E}$ 尺度矩阵意义下).

根据上述扩展了的最速下降方向的概念，梯度法、Newton 法以及本章介绍的拟 Newton 法都是在不同尺度矩阵意义下的最速下降法.

1. 梯度法($\boldsymbol{A}=\boldsymbol{E}$ 尺度矩阵意义下的最速下降法). 其迭代公式是

$$\begin{cases}\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\lambda_k\boldsymbol{p}_k,\\ \boldsymbol{p}_k=-\nabla f(\boldsymbol{x}_k),\\ \lambda_k:f(\boldsymbol{x}_k+\lambda_k\boldsymbol{p}_k)=\min\limits_{\lambda\geqslant 0}f(\boldsymbol{x}_k+\lambda\boldsymbol{p}_k).\end{cases}\qquad(6-2)$$

在式(6－2)中，$\boldsymbol{p}_k$ 为 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}_k$ 处的负梯度方向，亦即在 $\boldsymbol{E}$ 尺度矩阵意义下的最速下降方向，由于 $\boldsymbol{p}_k$ 构造简单，因此收敛速度慢(线性收敛)；

2. Newton 法($\boldsymbol{A}=\boldsymbol{G}_k$ 尺度矩阵意义下的最速下降法). 其迭代公

式是

$$\begin{cases} x_{k+1} = x_k + \lambda_k p_k, \\ p_k = -G_k^{-1} \nabla f(x_k), \\ \lambda_k : f(x_k + \lambda_k p_k) = \min_{\lambda \geq 0} f(x_k + \lambda p_k). \end{cases} \tag{6-3}$$

在式(6-3)中,Newton 矩阵 $G_k = \nabla^2 f(x_k) = \left[\frac{\partial^2 f(x_k)}{\partial x_i^{(k)} \partial x_j^{(k)}}\right]_{i,j=1,\cdots,n}$ 为正定对称阵,p_k 为 $f(x)$ 在 x_k 处的 Newton 方向,亦即在 $A = G_k$ 尺度矩阵意义下的最速下降方向.

Newton 法虽然收敛速度快,但由于存在 G_k 计算量大,逆阵 G_k^{-1} 计算困难,且在迭代过程中不能保证 G_k^{-1} 的正定性等等问题.在实际应用中可操作性较差.

3. 拟 Newton 法($A = H_k$ 尺度矩阵意义下的最速下降法).

现在我们要构造一种新的算法,它尽可能保持 Newton 法收敛速度快的优点,而避免 Newton 矩阵求逆的计算.为此引入新的迭代矩阵序列 $\{H_k\}$ 用以替代 $\{G_k^{-1}\}$,不仅要求 $H_k \approx G_k^{-1}$,且 H_k 易于计算.具 H_k 形式的拟 Newton 法的迭代公式是:

$$\begin{cases} x_{k+1} = x_k + \lambda_k p_k, \\ p_k = -H_k \nabla f(x_k), \\ \lambda_k : f(x_k + \lambda_k p_k) = \min_{\lambda \geq 0} f(x_k + \lambda p_k), \\ H_{k+1} = H_k + \Delta H_k, \quad k = 0,1,\cdots. \end{cases} \tag{6-4}$$

在式(6-4)中,p_k 为拟 Newton 方向,亦即在 H_k 尺度矩阵意义下的最速下降方向;H_{k+1} 为修正矩阵,ΔH_k 为修正项.要求 H_{k+1} 具有如下性质:

1. 满足拟 Newton 方程,即 $H_{k+1} y_k = s_k$,其中

$$s_k = x_{k+1} - x_k,\ y_k = g_{k+1} - g_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

容易验证 $G_{k+1} s_k = y_k$,即 $G_{k+1}^{-1} y_k = s_k$(G_{k+1}^{-1} 满足拟 Newton 方程).

因此,要求 H_{k+1} 满足拟 Newton 方程,也就是要求 H_{k+1} 逼近 G_{k+1}^{-1},拟 Newton 法也因此得名.

2. 具有遗传对称、正定功能.即是说若 H_k 对称正定,H_{k+1} 也应对称正定.因为 G_{k+1} 为对称阵,从而 G_{k+1}^{-1} 亦为对称阵,因此要求 H_{k+1} 必须是对称阵;要求 H_{k+1} 正定是为了保证 p_k 成为下降方向,从而保证迭代算法的下降性质.

修正项 ΔH_k 的构造与 H_k, s_k, y_k 有关,其结构由具体算法确定,通常可分为以下两种结构:

1. 对称秩 1 结构(SR1)

$$\Delta H_k = \sigma zz^T, \quad \mathrm{rank}(zz^T) = 1,$$

其中常数 σ,列向量 z 与 H_k, s_k, y_k 有关.

例 6.1　对称秩 1 校正公式(简称 SR1 校正公式)

SR1 校正公式首先由 Daicon(1959 年)提出来的,后来经 Broyden (1967 年)改进形成如下公式:

$$\begin{cases} H_{k+1} = H_k + \Delta H_k, \\ \Delta H_k = \dfrac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{< s_k - H_k y_k, y_k >}. \end{cases} \tag{6-5}$$

该公式满足拟 Newton 方程,且具遗传对称性功能,但并不能在一般情况下具有遗传正定性功能.但由于它对正定二次函数具有一些很好的性质,因而受到广泛的注意.

2. 对称秩 2 结构(SR2)

$$\Delta H_k = \sigma_1 zz^T + \sigma_2 ww^T,$$

其中常数 σ_1, σ_2 及列向量 z, w 仍与 H_k, s_k, y_k 有关.在 SR2 校正公式中,最受欢迎的是 DFP 公式.

例 6.2　DFP 公式

该公式首先是由 Davidon(1959 年)提出来的,后经 Fletcher 和 Powell (1963 年)做了改进,形成如下的公式:

$$\begin{cases} H_{k+1} = H_k + \Delta H_k, \\ \Delta H_k = \dfrac{s_k s_k^T}{s_k^T y_k} - \dfrac{(H_k y_k)(H_k y_k)^T}{y_k^T H_k y_k}. \end{cases} \tag{6-6}$$

DFP 公式不仅满足拟 Newton 方程,且具有遗传对称正定性功能,它还具有其他一些好的性质,成为梯度方法中最受欢迎的公式.

6.2　DFP 公式及 DFP 算法

6.2.1　DFP 公式及其基本性质

在 SR2 公式的一般形式中

$$H_{k+1} = H_k + \sigma_1 zz^T + \sigma_2 w_k w_k^T,$$

令

$$\sigma_1 = \frac{1}{s_k^T y_k}, \quad \sigma_2 = -\frac{1}{y_k^T H_k y_k}, \quad z_k = s_k, w_k = H_k y_k,$$

便得到著名的 DFP 公式(6-6)

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(H_k y_k)(H_k y_k)^T}{y_k^T H_k y_k}.$$

将式(6-6)嵌入式(6-4)中便得到完整的拟 Newton 法迭代公式

给定初始向量 x_0 及初始矩阵 $H_0=I$(对称正定),按如下公式迭代:

$$\begin{cases} x_{k+1} = x_k + \lambda_k p_k, \text{其中} \\ \quad p_k = -H_k \nabla f(x_k) = -H_k g_k,\ k=0,1,\cdots, \\ \quad \lambda_k := f(x_k+\lambda_k p_k) = \min\limits_{\lambda\geqslant 0} f(x_k+\lambda p_k), \\ H_{k+1} = H_k + \Delta H_k, \\ \Delta H_k = \dfrac{s_k s_k^{\mathrm T}}{s_k^{\mathrm T} y_k} - \dfrac{(H_k y_k)(H_k y_k)^{\mathrm T}}{y_k^{\mathrm T} H_k y_k}. \end{cases} \tag{6-7}$$

下面我们首先给出 DFP 公式(6-6)的一些基本性质,然后给出实现迭代公式(6-7)的 DFP 算法.

DFP 公式(6-6)具有如下性质:

性质 1 H_{k+1}满足拟 Newton 方程.

因为

$$\begin{aligned} H_{k+1}y_k &= \left(H_k + \frac{s_k s_k^{\mathrm T}}{s_k^{\mathrm T} y_k} - \frac{(H_k y_k)(H_k y_k)^{\mathrm T}}{y_k^{\mathrm T} H_k y_k}\right) y_k \\ &= H_k y_k + \frac{s_k(s_k^{\mathrm T} y_k)}{s_k^{\mathrm T} y_k} - \frac{(H_k y_k)(y_k^{\mathrm T} H_k y_k)}{y_k^{\mathrm T} H_k y_k} \\ &= H_k y_k + s_k - H_k y_k = s_k, \end{aligned}$$

由此得 $H_{k+1}y_k = s_k$,即 H_{k+1}满足拟 Newton 方程.

性质 2 H_{k+1}具有遗传对称、正定性功能.

在式(6-6)中,设若 H_k 对称,则由矩阵 $s_k s_k^{\mathrm T}$ 及$(H_k y_k)(H_k y_k)^{\mathrm T}$ 的对称性,可知 H_{k+1}为对称矩阵;若 H_k 正定,可以证明 H_{k+1}仍正定(证明见 6.2.2 定理 6.3).由性质 2 知,若给定的初始矩阵 H_0 对称正定,则公式(6-6)计算的任何 H_k 都是对称正定矩阵.

性质 3 公式(6-6)满足非正交条件,即对 $\forall k\geqslant 0, s_k^{\mathrm T} y_k \neq 0, y_k^{\mathrm T} H_k y_k \neq 0$.

因为 $x_{k+1}-x_k=\lambda_k p_k, p_k=-H_k g_k$,

所以

$$\begin{aligned} s_k^{\mathrm T} y_k &= (x_{k+1}-x_k)^{\mathrm T}(g_{k+1}-g_k) = (\lambda_k p_k)^{\mathrm T}(g_{k+1}-g_k) \\ &= \lambda_k p_k^{\mathrm T} g_{k+1} - \lambda_k p_k^{\mathrm T} g_k = -\lambda_k p_k^{\mathrm T} g_k \quad (p_k^{\mathrm T} g_{k+1} = g_{k+1}^{\mathrm T} p_k = 0) \\ &= \lambda_k (H_k g_k)^{\mathrm T} g_k = \lambda_k g_k^{\mathrm T} H_k g_k. \end{aligned}$$

又因为 $\lambda_k>0$,故对 $\forall g_k \neq 0$,由 $H_k>0$ 得到 $g_k^{\mathrm T} H_k g_k>0$,即 $s_k^{\mathrm T} y_k>0$.

由于 H_k 正定,即对 $\forall y_k\neq 0, y_k^{\mathrm T} H_k y_k>0$.

非正交条件又称为式(6-6)的存在条件,它保证矩阵迭代过程可以进行下去.

6.2.2 DFP 算法

在使用迭代公式(6－7)进行计算时，由于计算中舍入误差的影响，特别是在计算步长因子 λ_k 时，执行一维寻查不精确的影响，可能会破坏迭代矩阵 $\boldsymbol{H}_k$ 的正定性，从而导致算法失效，为保证 $\boldsymbol{H}_k$ 的正定性，可采取以下两条措施：

1. 一维寻查之后，函数值不下降时，重置 $\boldsymbol{H}_k$ 为单位矩阵，然后继续迭代.

2. 仿效共轭梯度法，迭代 $n+1$ 次后，重置初始点和迭代矩阵，即置 $\boldsymbol{x}_0=\boldsymbol{x}_{n+1}$，$\boldsymbol{H}_n=\boldsymbol{I}$ 以后重新迭代(参看框图 6－1).

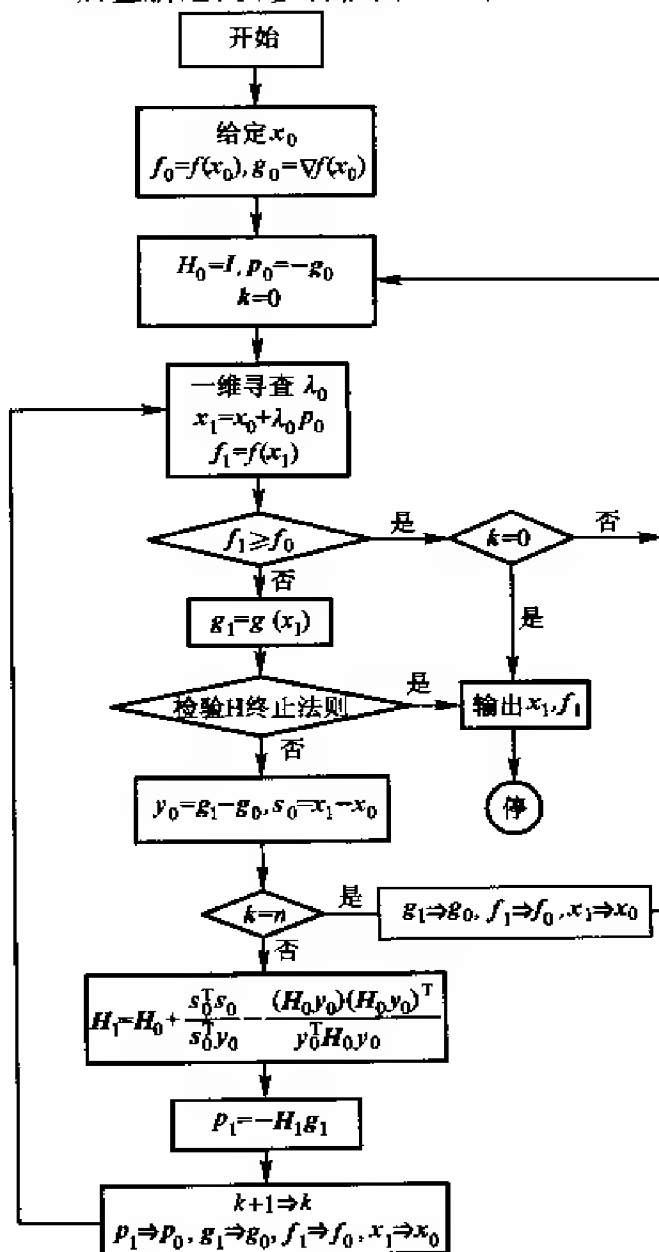


图 6－1

算法 6.1(DFP 算法)

设目标函数 $f(x), x \in \mathbf{R}^n$，H 终止法则中的误差限为 $\varepsilon_1, \varepsilon_2$ 和 ε_3.

① 给定初始点 x_0；计算 $f_0 = f(x_0), g_0 = \nabla f(x_0)$；

② 置 $H_0 = I, p_0 = -g_0, k = 0$；

③ 计算 $\lambda_k: f(x_k + \lambda_k p_k) = \min\limits_{\lambda \geq 0} f(x_k + \lambda p_k)$；

④ 计算 $x_{k+1} = x_k + \lambda_k p_k, f_{k+1} = f(x_{k+1})$；

⑤ 判别 $f_{k+1} \geq f_k$?

是，转⑥；否则转⑦；

⑥ 若 $k = 0$，则转⑬；否则转②；

⑦ 计算 $g_{k+1} = \nabla f(x_{k+1})$；

⑧ 判别 H - 终止法则是否满足?

是，则转⑬；否则转⑨；

⑨ 计算 $y_k = g_{k+1} - g_k, s_k = x_{k+1} - x_k$；

⑩ 若 $k = n$，则置 $x_0 = x_{k+1}, f_0 = f_{k+1}, g_0 = g_{k+1}$，转②；否则转⑪；

⑪ 计算 $H_{k+1} = H_k + \dfrac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} - \dfrac{(H_k y_k)(H_k y_k)^{\mathrm{T}}}{y_k^{\mathrm{T}} H_k y_k}$，

$$p_{k+1} = -H_{k+1} g_{k+1};$$

⑫ 置 $k = k+1, p_k = p_{k+1}, g_k = g_{k+1}, f_k = f_{k+1}, x_k = x_{k+1}$，转③；

⑬ 打印 x_{k+1}, f_{k+1}；

⑭ Stop.

例 6.3　用 DFP 算法求解　$\min\{x_1^2 + 4x_2^2\}$.

解　取 $x_0 = (1,1)^{\mathrm{T}}, H_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,

第一次迭代，DFP 算法与梯度法具有相同的迭代点 x_1，在第 2 章 2.3 中已作了计算，结果如下：

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \quad g(x) = \nabla(2x_1, 8x_2)^{\mathrm{T}},$$

$$x_0 = (1,1)^{\mathrm{T}}, \quad g_0 = (2,8)^{\mathrm{T}},$$

$$x_1 = (0.738\ 46, -0.046\ 16)^{\mathrm{T}},$$

$$g_1 = \nabla f(x_1) = (1.476\ 92, -0.369\ 23)^{\mathrm{T}},$$

以下用 DFP 算法作第二次迭代，

$$s_0 = x_1 - x_0 = (-0.261\ 54, -1.046\ 16)^{\mathrm{T}},$$

$$y_0 = g_1 - g_0 = (-0.523\ 08, -8.369\ 23)^{\mathrm{T}},$$

$$H_1 = H_0 + \frac{s_0 s_0^{\mathrm{T}}}{s_0^{\mathrm{T}} y_0} - \frac{(H_0 y_0)(H_0 y_0)^{\mathrm{T}}}{y_0^{\mathrm{T}} H_0 y_0},$$

其中

$$s_0^T y_0 = 8.89236,$$

$$y_0^T H_0 y_0 = y_0^T y_0 = 70.31762,$$

$$s_0 s_0^T = \begin{bmatrix} 0.06840 & 0.27361 \\ 0.27361 & 1.09445 \end{bmatrix}.$$

$$(H_0 y_0)(H_0 y_0)^T = H_0 y_0 y_0^T H_0^T = y_0 y_0^T = \begin{bmatrix} 0.27361 & 4.37778 \\ 4.37778 & 70.04401 \end{bmatrix},$$

所以

$$H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.00769 & 0.03077 \\ 0.03077 & 0.12308 \end{bmatrix} - \begin{bmatrix} 0.00389 & 0.06226 \\ 0.06226 & 0.99611 \end{bmatrix}$$

$$= \begin{bmatrix} 1.00380 & -0.03149 \\ -0.03149 & 0.12697 \end{bmatrix},$$

迭代方向 $p_1 = -H_1 g_1 = (-1.49416, 0.09340)^T$,

按正定二次函数一维寻查公式计算 λ_1:

$$\lambda_1 = \frac{p_1^T g_1}{g_1^T A g_1} = 0.49423,$$

所以 $x_2 = x_1 + \lambda_1 p_1 = (0.00000, 0.00000)^T$,

又因为 $g_2 = \nabla f(x_2) = (0,0)^T$,目标函数为正定二次函数,

所以 x_2 是本问题的全局极小点.

6.2.3 DFP 算法的二次收敛性质

定理 6.1 设目标函数 $f(x) = \frac{1}{2}x^T A x + b^T x + c$,其中 $A^T = A, A > 0$,对 $f(x)$ 施行算法 6.1. 如果:

(1) 初始矩阵 H_0 是对称正定的;

(2) 迭代点是互异的,

则(1)由算法 6-1 所生成的迭代方向 $\{p_i\}$ $(i = 0,1,\cdots,k+1)$ 是 A - 共轭向量系,即

$$p_i^T A p_j = 0, i \neq j; i,j = 0,1,\cdots,k+1;$$

(2) 对 $j = 0,1,\cdots,k$ 下式成立:$H_{k+1} A p_j = p_j$.

本定理中的两个结论可以用数学归纳法证明,由于证明过程中需要用较多的辅助等式,故证明从略.

定理 6.1 说明了 DFP 算法同时是一种共轭方向法,因此具有二次收敛性质. 对于正定 n 元二次函数,最多经过 n 次迭代就可达到极小点. 在例 6.3 中,目标函数为正定二元二次函数,从初始点 $x_0 = (1,1)^T$,经过两次迭代求得极小点 $x_2 = (0,0)^T$.

还要指出，当迭代一直进行到 n 次才结束时，必有 $H_n = A^{-1}$，即是说迭代矩阵 $H_1, H_2, \cdots, H_n, \cdots$ 是逐次逼近二次函数 $f(x)$ 的 Hession 逆阵的. 下面给出这一结论的证明.

引理 6.1　设 A 是 n 阶方阵，$v_1, v_2, \cdots, v_n$ 是线性无关的 n 维向量组，若对于 $j = 0, 1, \cdots, n$ 都有

$$Av_j = v_j,$$

则 A 是 n 阶单位阵，即 $A = I$.

证　据题设条件，对于 $\forall v \in \mathbf{R}^n$，总存在 $c_i (i = 1, \cdots, n)$，使得

$$v = \sum_{j=1}^{n} c_j v_j,$$

而且
$$Av = A\left(\sum_{j=1}^{n} c_j v_j\right) = \sum_{j=1}^{n} c_j A v_j = \sum_{j=1}^{n} c_j v_j = v = Iv,$$

特别地令
$$v = e_j = (0, \cdots, 0, 1, 0 \quad \cdots, 0)^{\mathrm{T}},$$

代入上式有
$$Ae_j = Ie_j, \qquad j = 1, 2, \cdots, n,$$

即
$$A_j = I_j, \qquad j = 1, 2, \cdots, n,$$

所以
$$A = I.$$

定理 6.2　若定理 6.1 的条件均满足，并且经过 n 次迭代达到极小点，则 $H_n = A^{-1}$.

证　在定理 6.1 的结论(2)中，令 $k = n - 1$ 有

$$H_n A p_j = p_j, \quad j = 0, 1, \cdots, n-1,$$

又由定理 6.1 的结论(1)知，$p_0, p_1, \cdots, p_{n-1}$ 是 A - 共轭向量系，因此该向量系线性无关，于是由引理 6.1 得到

$$H_n A = I.$$

所以
$$H_n = A^{-1}.$$

将例 6.3 继续做下去，就可以验证定理 6.2.

在例 6.3 中已经求出

$$H_1 = \begin{bmatrix} 1.003\ 80 & -0.031\ 49 \\ -0.031\ 49 & 0.126\ 97 \end{bmatrix},$$

$$x_1 = (0.738\ 46,\ -0.046\ 16)^{\mathrm{T}},$$

$$g_1 = (1.476\ 92,\ -0.369\ 23)^{\mathrm{T}}$$

$$x_2 = (0, 0)^{\mathrm{T}}, \quad g_2 = (0, 0)^{\mathrm{T}},$$

因此
$$s_1 = x_2 - x_1 = (-0.738\ 46, 0.046\ 16)^{\mathrm{T}},$$

$$y_1 = g_2 - g_1 = (-1.476\ 92, 0.369\ 23)^{\mathrm{T}},$$

$$H_2 = H_1 + \frac{s_1 s_1^{\mathrm{T}}}{s_1^{\mathrm{T}} y_1} - \frac{(H_1 y_1)(H_1 y_1)^{\mathrm{T}}}{y_1^{\mathrm{T}} H_1 y_1},$$

其中
$$s_1^T y_1 = 1.107\ 69,$$
$$y_1^T H_1 y_1 = 2.241\ 24,$$
$$s_1 s_1^T = \begin{bmatrix} 0.545\ 32 & -0.034\ 09 \\ -0.034\ 09 & 0.002\ 13 \end{bmatrix},$$
$$(H_1 y_1)(H_1 y_1)^T = \begin{bmatrix} -1.494\ 16 \\ 0.093\ 39 \end{bmatrix}(-1.494\ 16, 0.093\ 39)$$
$$= \begin{bmatrix} 0.273\ 61 & 4.377\ 78 \\ 4.377\ 78 & 70.044\ 01 \end{bmatrix},$$
所以
$$H_2 = \begin{bmatrix} 1.003\ 80 & -0.031\ 49 \\ -0.031\ 49 & 0.126\ 97 \end{bmatrix} + \begin{bmatrix} 0.492\ 31 & -0.030\ 78 \\ -0.030\ 78 & 0.001\ 92 \end{bmatrix}$$
$$- \begin{bmatrix} 0.996\ 10 & -0.062\ 26 \\ -0.062\ 26 & 0.003\ 89 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.125 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{8} \end{bmatrix},$$
因为
$$H_2 A = I, \quad AH_2 = I,$$
所以
$$H_2 = A^{-1}.$$

在 6.2.1 性质 2 中提出 H_k 具有遗传正定性功能,在下面的定理中得到证明.

定理 6.3 在 DFP 算法中,若初始矩阵是对称正定的,则 $\{H_k\}$ 中的每一个矩阵都是正定的.

证 用数学归纳法. 当 $k=0$ 时,H_0 是正定的.

假设若 $k=m$ 时,H_m 是正定的,以下证明 H_{m+1} 也是正定的. 即对于 $\forall z \in \mathbf{R}^n$,当 $z \neq 0$ 时,总有
$$z^T H_{m+1} z > 0, \tag{6-8}$$
据 DFP 公式(6-6)有
$$H_{m+1} = H_m + \frac{s_m s_m^T}{s_m^T y_m} - \frac{(H_m y_m)(H_m y_m)^T}{y_m^T H_m y_m},$$
将上式代入(6-8)中有
$$z^T H_{m+1} z = z^T H_m z + \frac{(z^T s_m)^2}{s_m^T y_m} - \frac{(z^T H_m y_m)^2}{y_m^T H_m y_m}. \tag{6-9}$$
据 6.2.1 的性质 3(非正交条件),在式(6-9)中,
$$y_m^T s_m > 0, \quad y_m^T H_m y_m > 0,$$
因此式(6-9)总是成立的.

据题设 H_m 对称正定,故存在对称正定矩阵 L_m 使得

$$H_m = L_m L_m,$$

此时 $$z^T H_m z = z^T L_m L_m z = (L_m z)^T (L_m z),$$

$$y_m^T H_m y_m = (L_m y_m)^T (L_m y_m),$$

$$z^T H_m y_m = (L_m z)^T (L_m y_m),$$

令 $$p = L_m z, \quad q = L_m y_m \tag{6-10}$$

得到 $$z^T H_m z = p^T p,\ y_m^T H_m y_m = q^T q,\ z^T H_m y_m = p^T q.$$

将上述结果代入式(6-9)中得到

$$z^T H_{m+1} z = \frac{(p^T p)(q^T q) - (p^T q)^2}{q^T q} + \frac{(z^T s_m)^2}{y_m^T s_m}, \tag{6-11}$$

分两种情形讨论：

(1) 当 $p = \lambda q, \lambda \neq 0$ 时，式(6-11)中的第一项为0，现证明第二项一定大于0. 事实上，据(6-10)，$z = \lambda y_m$，代入式(6-11)的第二项中便有

$$\frac{(z^T s_m)^2}{y_m^T s_m} = \frac{\lambda^2 (y_m^T s_m)^2}{y_m^T s_m} = \lambda^2 (y_m^T s_m) > 0.$$

(2) 当 $p \neq \lambda q, \lambda \neq 0$ 时，由 Cauchy-Schwarz 不等式，式(6-11)的第一项大于0，而第二项非负.

综上，对 $\forall z \neq 0$，$z^T H_{m+1} z > 0$，即 H_{m+1} 正定. 定理得证.

6.3 DFP 对偶公式及其等价形式

6.3.1 DFP 对偶公式及其基本性质

引理 6.2(Sherman-Morrison 公式) 设 $u, v \in \mathbf{R}^n$，A 为 n 阶方阵，且 $\det(A) \neq 0$，则当 $1 + v^T A^{-1} u \neq 0$ 时，成立下式：

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u}. \tag{6-12}$$

证 $$(A + uv^T)(A + uv^T)^{-1} = (A + uv^T)\left(A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u}\right)$$

$$= I + uv^T A^{-1} - \frac{uv^T A^{-1}}{1 + v^T A^{-1} u} - \frac{(uv^T A^{-1})(uv^T A^{-1})}{1 + v^T A^{-1} u}$$

$$= I + \frac{(uv^T A^{-1})(v^T A^{-1} u)}{1 + v^T A^{-1} u} - \frac{(uv^T A^{-1})(uv^T A^{-1})}{1 + v^T A^{-1} u},$$

而 $$(uv^T A^{-1})(v^T A^{-1} u) = u(v^T A^{-1} u) v^T A^{-1} = (uv^T A^{-1})(uv^T A^{-1}),$$

故 $$(A + uv^T)(A + uv^T)^{-1} = I.$$

同理可证 $$(A + uv^T)^{-1}(A + uv^T) = \left(A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u}\right)(A + uv^T)$$

$=I.$

定理 6.4　在 DFP 公式(6－6)中，令 $B_k=H_k^{-1}, B_{k+1}=H_{k+1}^{-1}$，则有

$$B_{k+1}=B_k+\left(\frac{1}{y_k^T s_k}+\frac{s_k^T B_k s_k}{(y_k^T s_k)^2}\right)y_k y_k^T-\frac{1}{y_k^T s_k}(y_k(B_k s_k)^T+(B_k s_k)y_k^T), \tag{6-13}$$

称式(6－13)为对偶 DFP 公式.

证　在 DFP 公式(6－6)中，令

$$T_k=H_k+\frac{s_k s_k^T}{s_k^T y_k}, \tag{6-14}$$

有

$$H_{k+1}=T_k-\frac{(H_k y_k)(H_k y_k)^T}{y_k^T H_k y_k}, \tag{6-15}$$

因为 H_k 正定及非正交条件 $s_k^T y_k>0$，所以 T_k 亦正定，必然满足非奇异条件 $\det(T_k)\neq 0$，所以应用引理 6.2 于公式(6－15)中，得到

$$H_{k+1}^{-1}=T_k^{-1}+\frac{T_k^{-1}(H_k y_k)(H_k y_k)^T T_k^{-1}}{y_k^T H_k y_k+(H_k y_k)^T T_k^{-1}(H_k y_k)}. \tag{6-16}$$

由于 H_k, T_k^{-1} 皆正定，式中第二项分母大于零，保证式(6－16)存在性.

再对式(6－14)应用引理 6.2 得到

$$T_k^{-1}=H_k^{-1}-\frac{H_k^{-1}s_k s_k^T H_k^{-1}}{s_k^T y_k+s_k^T H_k s_k},$$

由于 $s_k^T y_k>0, s_k^T H_k s_k>0$，故该式成立.

将 T_k^{-1} 代入(6－16)中，经整理可得(6－13).

对偶 DFP 公式(6－13)具有 DFP 公式相同的性质：

性质 1　B_{k+1} 满足拟 Newton 方程 $B_{k+1}s_k=y_k$.

事实上，用 s_k 右乘式(6-13)左右两端得

$$B_{k+1}s_k=B_k s_k+y_k+\frac{s_k^T B_k s_k}{y_k^T s_k}y_k-B_k s_k-\frac{s_k^T B_k s_k}{y_k^T s_k}y_k=y_k.$$

性质 2　B_{k+1} 具有遗传对称、正定性功能.

在式(6－13)中，如果 B_k 对称，则因为矩阵 $y_k y_k^T$ 对称，矩阵 $(B_k s_k)^T+(B_k s_k)y_k^T$ 对称，立即可知 B_{k+1} 亦为对称阵；若 B_k 正定，我们也可以证明 B_{k+1} 仍正定(类似于 6.2.2 中定理 6.3 的证明).

性质 3　公式(6－13)仍满足非正交条件.

6.3.2　DFP 对偶公式的几种等价形式

1. 对称形式

$$B_{k+1} = \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) B_k \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) + \frac{y_k y_k^T}{y_k^T s_k}, \qquad (6-17)$$

将式(6－13)展开：

$$\begin{aligned} B_{k+1} &= B_k + \frac{y_k y_k^T}{y_k^T s_k} + \frac{s_k^T B_k s_k}{(y_k^T s_k)^2} y_k y_k^T - \frac{y_k (B_k s_k)^T}{y_k^T s_k} - \frac{(B_k s_k) y_k^T}{y_k^T s_k} \\ &= B_k + \frac{y_k y_k^T}{y_k^T s_k} + \frac{y_k (s_k^T B_k s_k) y_k^T}{(y_k^T s_k)^2} - \frac{(B_k s_k) y_k^T}{y_k^T s_k} - \frac{y_k (B_k s_k)^T}{y_k^T s_k}, \end{aligned}$$

将上式的第 1、3、4、5 项合并为$\left(I - \frac{y_k y_k^T}{y_k^T s_k}\right) B_k \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right)$,

故得 $$B_{k+1} = \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) B_k \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) + \frac{y_k y_k^T}{y_k^T s_k}.$$

2. Powell 形式

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) y_k^T + y_k (y_k - B_k s_k)^T}{y_k^T s_k} - \frac{(y_k - B_k y_k)^T s_k}{(y_k^T s_k)^2} y_k y_k^T, \qquad (6-18)$$

将(6－13)适当变形

$$\begin{aligned} B_{k+1} &= B_k + \frac{y_k^T s_k + s_k^T B_k s_k}{(y_k^T s_k)^2} y_k y_k^T - \frac{1}{y_k^T s_k}\left(y_k (B_k s_k)^T + (B_k s_k) y_k^T\right) \\ &= B_k + \frac{-y_k^T s_k + s_k^T B_k s_k}{(y_k^T s_k)^2} y_k y_k^T + \frac{2 y_k^T s_k}{(y_k^T s_k)^2} y_k y_k^T - \frac{1}{y_k^T s_k}\left(y_k (B_k s_k)^T + (B_k s_k) y_k^T\right) \\ &= B_k - \frac{y_k^T s_k - s_k^T B_k s_k}{(y_k^T s_k)^2} y_k y_k^T + \frac{2 y_k y_k^T - y_k (B_k s_k)^T - (B_k s_k) y_k^T}{y_k^T s_k} \\ &= B_k - \frac{(y_k - B_k s_k)^T s_k}{(y_k^T s_k)^2} y_k y_k^T + \frac{y_k y_k^T - (B_k s_k) y_k^T + y_k y_k^T - y_k (B_k s_k)^T}{y_k^T s_k} \\ &= B_k + \frac{(y_k - B_k s_k) y_k^T + y_k (y_k - B_k s_k)^T}{y_k^T s_k} - \frac{(y_k - B_k s_k)^T s_k}{(y_k^T s_k)^2} y_k y_k^T, \end{aligned}$$

由此得到式(6－18).

以上两种形式具有对称，便于记忆的优点，但在 DFP 对偶算法设计中，下面的形式更为实用.

3. 对称秩 2 的形式

$$B_{k+1} = B_k + \frac{1}{\mu_k^2 \delta_k^2} t_k t_k^T - \mu_k^2 g_k g_k^T, \text{其中}$$

$$\begin{cases} t_k = (\mu_k^2\delta_k - 1)g_k + g_{k+1}, \\ \mu_k^2 = \dfrac{\lambda_k}{\delta_k - \lambda_k\eta_k}, \\ \delta_k = y_k^{\mathrm{T}}p_k, \\ \eta_k = g_k^{\mathrm{T}}p_k. \end{cases} \tag{6-19}$$

在式(6-13)中,$s_k = x_{k+1} - x_k = \lambda_k p_k, y_k = g_{k+1} - g_k$,

$$B_k s_k = \lambda_k B_k p_k = -\lambda_k g_k \quad (p_k = -H_k g_k, B_k p_k = -g_k),$$

将 $s_k, y_k, B_k s_k$ 代入(6-13)中,有

$$B_{k+1} = B_k + \left(\frac{1}{\lambda_k y_k^{\mathrm{T}} p_k} - \frac{g_k^{\mathrm{T}} p_k}{(y_k^{\mathrm{T}} p_k)^2}\right)(g_{k+1} - g_k)(g_{k+1} - g_k)^{\mathrm{T}}$$

$$+ \frac{1}{y_k^{\mathrm{T}} p_k}((g_{k+1} - g_k)g_k^{\mathrm{T}} + g_k(g_{k+1} - g_k)^{\mathrm{T}}).$$

令 $y_k^{\mathrm{T}} p_k = \delta_k, g_k^{\mathrm{T}} p_k = \eta_k$,有

$$B_{k+1} = B_k + \frac{\delta_k - \lambda_k\eta_k}{\lambda_k\delta_k^2}(g_{k+1}g_{k+1}^{\mathrm{T}} + g_k g_k^{\mathrm{T}})$$

$$+ \left(\frac{1}{\delta_k} - \frac{\delta_k - \lambda_k\eta_k}{\lambda_k\delta_k^2}\right)(g_{k+1}g_k^{\mathrm{T}} + g_k g_{k+1}^{\mathrm{T}}) - \frac{2}{\delta_k}g_k g_k^{\mathrm{T}}.$$

因为 $\delta_k - \lambda_k\eta_k = y_k^{\mathrm{T}} p_k - \lambda_k g_k^{\mathrm{T}} p_k = (g_{k+1} - g_k)^{\mathrm{T}} p_k - \lambda_k g_k^{\mathrm{T}} p_k$

$$= -g_k^{\mathrm{T}} p_k - \lambda_k g_k^{\mathrm{T}} p_k = (1 + \lambda_k) g_k^{\mathrm{T}} H_k g_k > 0,$$

于是令 $\mu_k^2 = \dfrac{\lambda_k}{\delta_k - \lambda_k\eta_k}$,则有

$$B_{k+1} = B_k + \frac{1}{\mu_k^2\delta_k^2}(g_{k+1}g_{k+1}^{\mathrm{T}} + g_k g_k^{\mathrm{T}})$$

$$+ \left(\frac{1}{\delta_k} - \frac{\delta_k - \lambda_k\eta_k}{\mu_k^2\delta_k^2}\right)(g_{k+1}g_k^{\mathrm{T}} + g_k g_{k+1}^{\mathrm{T}}) - \frac{2}{\delta_k}g_k g_k^{\mathrm{T}}$$

$$= B_k + \frac{1}{\mu_k^2\delta_k^2}((\mu_k^2\delta_k - 1)g_k + g_{k+1})((\mu_k^2\delta_k - 1)g_k + g_{k+1})^{\mathrm{T}} - \mu_k^{\mathrm{T}} g_k g_k^{\mathrm{T}},$$

再令 $t_k = (\mu_k^2\delta_k - 1)g_k + g_{k+1}$,得到

$$B_{k+1} = B_k + \frac{1}{\mu_k^2\delta_k^2}t_k t_k^{\mathrm{T}} - \mu_k^2 g_k g_k^{\mathrm{T}}.$$

6.4 DFP 对偶算法

给定初始向量 x_0 及初始矩阵 B_0(对称正定),DFP 对偶算法如迭代

公式(6－20)所示：

$$\begin{cases} x_{k+1} = x_k + \lambda_k p_k, \\ p_k: B_k p_k = -g_k, \\ \lambda_k: \min\limits_{\lambda \geq 0} f(x_k + \lambda p_k), \\ B_{k+1} = B_k + \Delta B_k \quad k = 0,1,\cdots, \\ \quad 其中 \Delta B_k 由(6-13)或(6-19)确定. \end{cases} \tag{6-20}$$

在上述算法中，拟 Newton 方向 p_k 是通过解线性方程组 $B_k p_k = -g_k$ 来确定的. 鉴于 B_k 为对称正定矩阵，通常采用 LDL^T 分解方法解线性方程组以保证数值稳定性. 即将 B_k 分解为 $B_k = L_k D_k L_k^T$，通过解两组三角方程

$$\begin{cases} L_k v_k = -g_k, \\ L_k^T p_k = D_k^{-1} v_k \end{cases}$$

确定 p_k.

为了提高算法的效率，我们不希望每次迭代都进行 LDL^T 分解，而是导出 L_k, D_k 的迭代公式，直接由 L_k, D_k 计算 L_{k+1}, D_{k+1}.

具体地讲，即若已知 $B_k = L_k D_k L_k^T$，由对称秩 2 公式

$$B_{k+1} = B_k + \sigma_1 zz^T + \sigma_2 ww^T$$

求出 B_{k+1} 的 Cholesky 因子 L_{k+1}, D_{k+1}，使

$$B_{k+1} = L_{k+1} D_{k+1} L_{k+1}^T.$$

6.4.1　修正矩阵 B_{k+1} 的 LDL^T 分解

1. 问题的提法

对于对称正定矩阵 B，可以实现 LDL^T 分解，即 $B = LDL^T$，其中 L 为单位下三角阵，D 为对角阵.

$$L = \begin{bmatrix} 1 & & & & \\ l_{21} & 1 & & & \\ l_{31} & l_{32} & 1 & & \\ \vdots & \vdots & & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{n,n-1} & 1 \end{bmatrix},$$

$$D = \mathrm{diag}(d_1, d_2, \cdots, d_n)$$

$$d_i > 0 (i = 1, \cdots, n).$$

设 $B^* = B + \sigma zz^T$，其中 σ 为常数，$\mathrm{rank}(zz^T) = 1$，求 B^* 的 LDL^T 分解，即求 L^*, D^* 使 $B^* = B + \sigma zz^T = L^* D^* L^{*T}$，其中 L^* 仍为单位下三角阵，D^* 仍为对角阵.

在上述问题中，我们不希望直接对 B^* 进行 $L^*D^*L^{*T}$ 分解，而是希望通过对 B 进行 LDL^T 分解后，以 L 阵和 D 阵的元素去递推 L^* 阵和 D^* 阵的元素.

设 $\tilde{v}=(\tilde{v}_1,\cdots,\tilde{v}_n)^T$ 是三角方程 $L\tilde{v}=z$ 的解，则有

$$B^* = B + \sigma zz^T = L(D+\sigma\tilde{v}\tilde{v}^T)L^T,$$

因为 D 对称正定，$\sigma>0$，故 $D+\sigma\tilde{v}\tilde{v}^T$ 亦对称正定，若对 $D+\sigma\tilde{v}\tilde{v}^T$ 进行 LDL^T 分解，即

$$D+\sigma\tilde{v}\tilde{v}^T = \tilde{L}\tilde{D}\tilde{L}^T,$$

于是
$$B^* = (L\tilde{L})\tilde{D}(\tilde{L}^TL^T) = (L\tilde{L})\tilde{D}(L\tilde{L})^T.$$

令 $L^*=L\tilde{L}$, $D^*=\tilde{D}$ 得到 $B^*=L^*D^*L^{*T}$，显然 $L^*=L\tilde{L}$ 仍为单位下三角阵；而 $D^*=\tilde{D}$ 仍为正定对角阵. 并且计算矩阵 L^* 和 D^* 转化为计算 $\tilde{L}$ 和 $\tilde{D}$.

2. 分解 $D+\sigma\tilde{v}\tilde{v}^T=\tilde{L}\tilde{D}\tilde{L}^T$ 的计算公式，记

$$\tilde{D} = \mathrm{diag}(\tilde{d}_1,\tilde{d}_2,\cdots,\tilde{d}_n),\quad \tilde{L} = (\tilde{l}_1,\cdots,\tilde{l}_n),$$

其中 $\tilde{l}_i=(\tilde{l}_{1i},\cdots,\tilde{l}_{ni})^T$, $i=1,2,\cdots,n$.

记 $\tilde{l}_j^{(s)}$ 为 $\tilde{l}_j$ 的后 $n-s+1$ 个元素组成的 $n-s+1$ 维向量($s,j=1,2,\cdots,n$). 例如：

$$\tilde{l}_1^{(2)} = (\tilde{l}_{21},\tilde{l}_{31},\cdots,\tilde{l}_{n1})^T,\quad n-2+1=n-1 \text{ 个元素;}$$

$$\tilde{l}_2^{(3)} = (\tilde{l}_{32},\tilde{l}_{42},\cdots,\tilde{l}_{n2})^T,\quad n-3+1=n-2 \text{ 个元素;}$$

$$\tilde{l}_3^{(4)} = (\tilde{l}_{43},\tilde{l}_{53},\cdots,\tilde{l}_{n3})^T,\quad n-4+1=n-3 \text{ 个元素;}$$

…………

$$\tilde{l}_{n-1}^{(n)} = \tilde{l}_{n,n-1},\qquad n-n+1=1 \text{ 个元素.}$$

将 $\tilde{l}_1^{(2)},\tilde{l}_2^{(3)},\cdots,\tilde{l}_{n-1}^{(n)}$ 按列排列起来：

$$\begin{matrix} \tilde{l}_{21} & & & \\ \tilde{l}_{31} & \tilde{l}_{32} & & \\ \vdots & \vdots & \ddots & \\ \tilde{l}_{n1} & \tilde{l}_{n2} & \cdots & \tilde{l}_{n,n-1} \end{matrix}$$

恰好组成 $\tilde{L}$ 的下三角部分(除去主对角元素 1),又记 $\tilde{v}^{(i)}$ 为 $\tilde{v}$ 的后 $n-i+1$ 个元素组成的向量.

形如

$$\tilde{v}^{(1)}=(\tilde{v}_1,\tilde{v}_2,\cdots,\tilde{v}_n)^{\mathrm{T}},$$
$$\tilde{v}^{(2)}=(\tilde{v}_2,\tilde{v}_3,\cdots,\tilde{v}_n)^{\mathrm{T}},$$
$$\cdots\cdots\cdots\cdots$$
$$\tilde{v}^{(n)}=\tilde{v}_n,$$

则可以推导出计算 $\tilde{L},\tilde{D}$ 的元素的计算公式:

$$\begin{cases}\tilde{d}_1=d_1+\sigma\tilde{v}_1^2,\\ \tilde{d}_i=d_i+\tilde{v}_i^2\left(\sigma-\sum\limits_{k=1}^{i-1}\tilde{\beta}_k^2\tilde{d}_k\right), \quad i=2,3,\cdots,n,\\ \tilde{\beta}_1=\dfrac{\sigma\tilde{v}_1}{\tilde{d}_1},\\ \tilde{\beta}_i=\dfrac{\left(\sigma-\sum\limits_{k=1}^{i-1}\tilde{\beta}_k^2\tilde{d}_k\right)\tilde{v}_i}{\tilde{d}_i}, \quad i=2,3,\cdots,n,\\ \tilde{l}_i^{(i+1)}=\tilde{\beta}_i\tilde{v}^{(i+1)}, \quad i=1,2,\cdots,n-1.\end{cases} \tag{6-21}$$

公式(6-21)的计算流程图示如下:

已知 $d_1,d_2,\cdots,d_n;\tilde{v}_1,\tilde{v}_2,\cdots,\tilde{v}_n;\sigma$,按箭头所示方向依次计算 $\tilde{d}_i$,$\tilde{\beta}_i$,$\tilde{l}_i^{(i+1)}$.

$$\tilde{d}_1\to\tilde{\beta}_1\to\tilde{l}_1^{(2)}=\tilde{\beta}_1\tilde{v}^{(2)}=\tilde{\beta}_1(\tilde{v}_2,\tilde{v}_3,\cdots,\tilde{v}_n)^{\mathrm{T}},$$
$$\tilde{d}_2\to\tilde{\beta}_2\to\tilde{l}_2^{(3)}=\tilde{\beta}_2\tilde{v}^{(3)}=\tilde{\beta}_2(\tilde{v}_3,\tilde{v}_4,\cdots,\tilde{v}_n)^{\mathrm{T}},$$
$$\cdots\cdots\cdots\cdots$$
$$\tilde{d}_{n-1}\to\tilde{\beta}_{n-1}\to\tilde{l}_{n-1}^{(n)}=\tilde{\beta}_{n-1}\tilde{v}^{(n)}=\beta_{n-1}\tilde{v}_n,$$
$$\tilde{d}_n.$$

3. $B+\sigma zz^{\mathrm{T}}$ 的分解算法

已知 $B=LDL^{\mathrm{T}}$ 且 $B^*=B+\sigma zz^{\mathrm{T}}$,求 $B^*=L^*DL^{*\mathrm{T}}$,

给定 $L,D,z,\sigma>0$,

算法 6.2

① 解方程组 $L\tilde{v}=z$ 得 $v=(\tilde{v}_1,\cdots,\tilde{v}_n)^{\mathrm{T}}$;

② 置 $i=1$;

③ 计算 $\begin{cases} \tilde{d}_i = d_i + \sigma v_i^2 & (i = 1); \\ \tilde{d}_i = d_i + v_i^2\left(\sigma - \sum_{j=1}^{i-1}\tilde{\beta}_j^2\tilde{d}_j\right) & (i > 1). \end{cases}$

④ 检查 $i < n$?

是,计算 $\begin{cases} \tilde{\beta}_i = \dfrac{\sigma\tilde{v}_i}{\tilde{d}_i} & (i = 1), \\ \tilde{\beta}_i = \dfrac{\left(\sigma - \sum_{j=1}^{i-1}\tilde{\beta}_j^2\tilde{d}_j\right)\tilde{v}_i}{\tilde{d}_i} & (i > 1), \end{cases}$ 转⑤;

否,转⑦;

⑤ 计算 $\tilde{l}_i^{(i+1)} = \beta_i\tilde{v}^{(i+1)} = \beta_i(\tilde{v}_{i+1}, \tilde{v}_{i+2}, \cdots, \tilde{v}_n)$;

⑥ $i+1 \Rightarrow i$ 转③;

⑦ 形成 $\tilde{L}$ 阵及 $\tilde{D}$ 阵;

⑧ 令 $L^* = L\tilde{L}, D^* = \tilde{D}$.

4. $B + \sigma_1 zz^T + \sigma_2 uu^T$ 的分解算法.

已知 $B_k = L_k D_k L_k^T$,且 $B_{k+1} = B_k + \sigma_1 t_k t_k^T + \sigma_2 g_k g_k^T$($\sigma_1, \sigma_2, t_k, g_k$ 为已知),求 $B^* = L^* D L^{*T}$.

算法 6.3(子程序)

① 置 $m = 1$;

② 赋初值 $\sigma = \sigma_1, L = L_k, D = D_k, y = t_k$;

③ 解 $L\tilde{v} = y$ 得 $\tilde{v} = (\tilde{v}_1, \tilde{v}_2, \cdots, \tilde{v}_n)^T$;

④ 置 $i = 1$;

⑤ 计算 $\begin{cases} \tilde{d}_i = d_i + \sigma v_i^2 & (i = 1); \\ \tilde{d}_i = d_i + v_i^2\left(\sigma - \sum_{j=1}^{i-1}\tilde{\beta}_j^2\tilde{d}_j\right) & (i > 1). \end{cases}$

⑥ 检查 $i < n$?

是,计算 $\begin{cases} \tilde{\beta}_i = \dfrac{\sigma\tilde{v}_i}{\tilde{d}_i} & (i = 1), \\ \tilde{\beta}_i = \dfrac{\left(\sigma - \sum_{j=1}^{i-1}\tilde{\beta}_j^2\tilde{d}_j\right)\tilde{v}_i}{\tilde{d}_i} & (i > 1), \end{cases}$ 转⑦;

否，转⑨；

⑦ 计算 $\tilde{l}_i^{(i+1)}=\beta_i\tilde{v}^{(i+1)}$；

⑧ $i+1\Rightarrow i$ 转⑤；

⑨ 形成 $\tilde{L},\tilde{D}$；

⑩ 计算 $L^*=L\tilde{L},D^*=\tilde{D}$；

⑪ 检查 $m=2$？

是，令 $L=L^*,D=D^*$，返回主程序（见算法 6.3 的⑮步）.

否，$m+1\Rightarrow m,\sigma_2\Rightarrow\sigma_1$，令 $L_k=L^*,D_k=D^*,t_k=g_k$ 转②.

6.4.2 带 LDL^T 分解的 DFP 对偶算法

算法 6.4

已知 $f(x)$ 及其梯度 $g(x)$，$x\in\mathbf{R}^n$，H－终止法则中的误差限 $\varepsilon_1,\varepsilon_2$ 和 ε_3.

① 给定 x_0，计算 $f_0=f(x_0),g_0=g(x_0)$；

② 置 $B_0=I$，并对 B_0 进行 $L_0D_0L_0^T$ 分解，即置 $L_0=I,D_0=I$；

③ 置 $k=0$；

④ 解 $B_kp_k=-g_k$，即由三角方程组$\begin{cases}L_kv_k=-g_k,\\L_k^Tp_k=D^{-1}v_k,\end{cases}$求解 p_k；

⑤ 求 $\lambda_k:\min\limits_{\lambda\geq0}f(x_k+\lambda p_k)$；

⑥ 计算 $x_{k+1}=x_k+\lambda_kp_k,f_{k+1}=f(x_{k+1})$；

⑦ 判别 $f_{k+1}\geq f_k$？

是，转⑧，

否，转⑨；

⑧ 若 $k=0$ 则转⑯；否则转②；

⑨ 计算 $g_{k+1}=g(x_{k+1})$；

⑩ 判别 H－终止法则是否满足？

若满足，转⑯；否则转⑪；

⑪ 计算 $y_k=g_{k+1}-g_k,s_k=x_{k+1}-x_k$；

⑫ 判别 $k=n$？

是，$x_{k+1}\Rightarrow x_0,f_{k+1}\Rightarrow f_0,g_{k+1}\Rightarrow g_0$，转②；

否则转⑬；

⑬ 计算 $\delta_k=y_k^Tp_k,\eta_k=g_k^Tp_k,\mu_k=\sqrt{\dfrac{\lambda_k}{\delta_k-\lambda_k\eta_k}},t_k=(\mu_k^2\delta_k-1)g_k$

$+g_{k+1}$;

⑭ 计算 $\sigma_1 = \dfrac{1}{\mu_k^2 \sigma_k^2}, \sigma_2 = -\mu_k^2$;

调用子程序(算法 6.2)对 $B_{k+1} = B_k + \sigma_1 t_k t_k^T + \sigma_2 g_k g_k^T$,进行 $L^* D^* L^{*T}$ 分解,并令 $L_{k+1} = L, D_{k+1} = D$;

⑮ $k+1 \Rightarrow k$ 转④;

⑯ 打印 x_{k+1}, f_{k+1};

⑰ Stop.

例 6.4 用对偶 DFP 算法求解 $\min\{x_1^2 + 4x_2^2\}$,取 $x_0 = (1,1)^T$.

解 取 $B_0 = I$,在第一次迭代,DFP 对偶算法与梯度法相同,其计算结果如下:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}, \quad g(x) = \nabla(2x_1, 8x_2)^T,$$

$$x_0 = (1,1)^T, \quad g_0 = (2,8)^T,$$

$$x_1 = (0.738\ 46, -0.046\ 16)^T,$$

$$g_1 = \nabla f(x_1) = (1.476\ 92, -0.369\ 23)^T,$$

用 DFP 对偶公式(6-13)作第二次迭代,

$$s_0 = x_1 - x_0 = (-0.261\ 54, -1.046\ 16)^T,$$

$$y_0 = g_1 - g_0 = (-0.523\ 08, -8.369\ 23)^T,$$

$$B_1 = B_0 + \left(\frac{1}{y_0^T s_0} + \frac{s_0^T B_0 s_0}{(y_0^T s_0)^2}\right) y_0 y_0^T - \frac{1}{y_0^T s_0}\left(y_0 (B_0 s_0)^T + (B_0 s_0) y_0^T\right).$$

将以下各式代入上式的右端得 B_1,

$$y_0^T s_0 = 8.892\ 36,$$

$$s_0^T B_0 s_0 = s_0^T s_0 = 1.162\ 85,$$

$$y_0 (B_0 s_0)^T = y_0 s_0^T = \begin{bmatrix} 0.136\ 81 & 0.547\ 23 \\ 2.188\ 89 & 8.755\ 56 \end{bmatrix},$$

$$(B_0 s_0) y_0^T = s_0 y_0^T = \begin{bmatrix} 0.136\ 81 & 2.188\ 89 \\ 0.547\ 23 & 8.755\ 56 \end{bmatrix},$$

$$y_0 s_0^T + s_0 y_0^T = \begin{bmatrix} 0.273\ 62 & 2.736\ 12 \\ 2.736\ 12 & 17.511\ 12 \end{bmatrix},$$

求出
$$B_1 = \begin{bmatrix} 1.004\ 0 & 0.248\ 99 \\ 0.248\ 99 & 7.937\ 70 \end{bmatrix},$$

解方程组 $B_1 p_1 = -g_1$,求拟 Newton 方向 $p_1 = (-1.494\ 16, 0.093\ 39)^T$,

求迭代步长因子
$$\lambda_1 = -\frac{p_1^T g_1}{p_1^T A p_1} = 0.494\ 23,$$

$$x_2 = x_1 + \lambda_1 p_1 = (0.000\,00,\ -0.000\,00)^T,$$

因为
$$g_2 = g(x_2) = (0,0)^T,$$

所以　x_2 是本问题的极小点.

第7章　直接搜索方法

前面几章介绍的无约束最优化算法都需要计算目标函数$f(x)$的梯度，但在实际中的很多目标函数要解析地计算它的偏导数常常是很困难，甚至是不可能的，例如：$f(x)$是由实测数据给出，不具有解析式；$f(x)$的解析式过于复杂，不便计算其偏导数；$f(x)$非公式表达但具有解析性，其偏导数只能用数值微分公式近似计算；无论$f(x)$是否由公式给出，计算偏导数存在病态问题等等.

综上所述，迫切需要引入一种直接搜索算法，实现这种算法不需要计算偏导数，甚至不必要求偏导数一定存在，只需要计算$f(x)$的值就够了.

常用的直接搜索方法包括：

1. 坐标轮换法（Cyclic coodinate）和步长加速法（Hooke-Jeeves 方法）. 这两种方法适用于低维问题；
2. 单纯形替换法（Neldes-Mead 方法）；
3. 转轴法（Rosenbrock 方法）以及 D. S. C 方法（Davies, Swahn, campey），后者即是采用精确一维寻查的 Rosenbrock 方法；
4. 方向加速法（Powell 方法）.

本章只介绍单纯形替换法和方向加速法.

7.1　单纯形替换法

7.1.1　$\mathbf{R}^n$ 中的单纯形

在 1.1 中的定义 1.3，我们介绍了 $\mathbf{R}^n$ 中的单纯形，它是具有 $n+1$ 个顶点的凸多面体. 若各个棱相等则称为正规单纯形. 例如在 $\mathbf{R}^2$ 中，三角形是单纯形，等边三角形则是正规单纯形.

我们把 $\mathbf{R}^n$ 中的单纯形记为：$\boldsymbol{H}(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_{n+1})$，其中 $\boldsymbol{v}_i(i=1,2,\cdots,n+1)$ 表示 $n+1$ 个顶点的位置向量；用$\overline{\boldsymbol{v}_i\boldsymbol{v}_j}(i=2,3,\cdots,n+1;j=1,2,\cdots,i-1)$ 表示 $\boldsymbol{H}(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_{n+1})$ 的$\dfrac{n(n+1)}{2}$个棱；用 $\boldsymbol{v}_{ij}(i=2,2,\cdots,n+1;j=1,2,\cdots,i-1)$ 表示$\overline{\boldsymbol{v}_i\boldsymbol{v}_j}$的中点.

将$\dfrac{n(n+1)}{2}$个中点排列成下表（见表 7－1）.

表 7－1

$$\begin{matrix} v_{21} & & & \\ v_{31} & v_{32} & & \\ \vdots & \vdots & \ddots & \\ v_{n+1,1} & v_{n+1,2} & \cdots & v_{n+1,n} \end{matrix}$$

$v_{ii}=v_i$ 表示中点退化为顶点.

为计算方便之见，单纯形的各个顶点及棱的中点通常用斜坐标系表示：

$v_1=(0,\cdots,0)^{\mathrm{T}}$ 表示过原点的顶点；

$v_{1j}=\frac{1}{2}(0,\cdots,0,\overset{(j-1)}{1},0,\cdots,0)^{\mathrm{T}}$ 表示过原点的棱的中点；

$v_{ij}=\frac{1}{2}(0,\cdots,0,\overset{(i-1)}{1},0,\cdots,0,\overset{(j-1)}{1},0,\cdots,0)^{\mathrm{T}}$ 表示不过原点的棱的中点；

$v_{ii}=v_i=\frac{1}{2}(0,\cdots,0,\overset{(i-1)}{1},0,\cdots,0)^{\mathrm{T}}$ 表示不过原点的顶点.

例如 $\mathbf{R}^2$ 中的单纯形 $H(v_1,v_2,v_3)$（图 7－1），

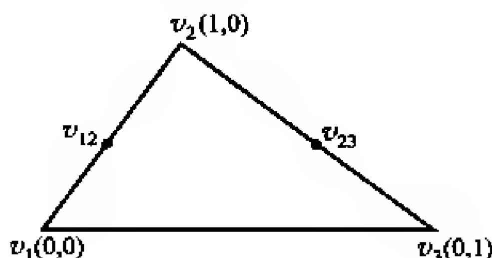


图 7－1

取 $v_1=(0,0)^{\mathrm{T}}$，　　$v_{12}=\frac{1}{2}(1,0)^{\mathrm{T}}$，

$v_{13}=\frac{1}{2}(0,1)^{\mathrm{T}}$，　　$v_{23}=\frac{1}{2}(1,1)^{\mathrm{T}}$，

$v_2=v_{22}=\frac{1}{2}(2,0)^{\mathrm{T}}=(1,0)^{\mathrm{T}}$，

$v_3=v_{33}=\frac{1}{2}(0,2)^{\mathrm{T}}=(0,1)^{\mathrm{T}}$.

下面我们介绍两种在 $\mathbf{R}^n$ 中构造一个单纯形的方法：

1. 构造正规单纯形（棱长相等）

给定正规单纯形的一个顶点 $x_0\in\mathbf{R}^n$ 及边长 s，则 $H(v_1,v_2,\cdots,v_{n+1})$ 的 $n+1$ 个顶点，按公式（7－1）计算：

$$
\begin{cases}
v_1 = x_0, \\
v_2 = (p, q, \cdots, q)^T + x_0, \\
v_3 = (q, p, q, \cdots, q)^T + x_0. \\
\cdots\cdots\cdots\cdots \\
v_i = (q, \cdots, q, \overset{(i-1)}{p}, q, \cdots, q)^T + x_0, \\
\cdots\cdots\cdots\cdots \\
v_{n+1} = (q, \cdots, q, \overset{(n)}{p})^T + x_0,
\end{cases}
\tag{7-1}
$$

其中 $p = \frac{s}{n\sqrt{2}}[\sqrt{n+1} + (n-1)], q = \frac{s}{n\sqrt{2}}(\sqrt{n+1} - 1).$

例如 $n=2$，取 $x^{(0)} = (x_1^{(0)}, x_2^{(0)})^T$，

$$
\begin{cases}
v_1 = (x_1^{(0)}, x_2^{(0)})^T, \\
v_2 = (p + x_1^{(0)}, q + x_2^{(0)})^T, \\
v_3 = (q + x_1^{(0)}, p + x_2^{(0)})^T,
\end{cases}
$$

其中
$$
p = \frac{s}{2\sqrt{2}}(\sqrt{3} + 1), \quad q = \frac{s}{2\sqrt{2}}(\sqrt{3} - 1).
$$

现在证明由式(7－1)所确定的单纯形 $H(v_1, v_2, \cdots, v_{n+1})$ 是正规单纯形.

首先将式(7－1)改写成如下公式

$$
\begin{cases}
v_1 = x_0, \\
v_i = x_0 + z_i, i = 2, \cdots, n+1,
\end{cases}
\tag{7-2}
$$

其中 n 维向量

$$
z_i = (q, \cdots, q, \overset{(i-1)}{p}, q, \cdots, q)^T.
$$

首先，对 $i = 2, 3, \cdots, n+1$ 计算

$$
\| v_i - v_1 \|_2^2 = \| x_0 + z_i - x_0 \|_2^2 = \| z_i \|_2^2 = p^2 + (n-1)q^2 = s^2.
$$

其次，对 $i, j = 2, 3, \cdots, n+1 (i \neq j)$ 计算

$$
\| v_i - v_j \|_2^2 = \| z_i - z_j \|_2^2 = \| z_i \|_2^2 - 2z_i^T z_j + \| z_j \|_2^2
$$

$$= 2s^2 - 2z_i^T z_j = 2s^2 - 2(2pq + (n-2)q^2) = s^2,$$

所以 $\boldsymbol{H}(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{n+1})$ 各棱相等，即为正规单纯形.

2. 构造非正规单纯形（棱长不全等）

设 s 为 $\boldsymbol{H}(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{n+1})$ 的某一边长，$n+1$ 个顶点按下式计算：

$$\begin{cases} \boldsymbol{v}_1 = \boldsymbol{x}_0, \\ \boldsymbol{v}_i = \boldsymbol{x}_0 + s\boldsymbol{e}_{i-1} \ (i = 2, \cdots, n+1), \end{cases} \tag{7-3}$$

其中
$$\boldsymbol{e}_{i-1} = (0, \cdots, 0, \overset{(i-1)}{1}, 0, \cdots, 0)^T.$$

当 $n=2$ 时，

$$\begin{cases} \boldsymbol{v}_1 = \boldsymbol{x}_0, \\ \boldsymbol{v}_2 = \boldsymbol{x}_0 + s\boldsymbol{e}_1, \\ \boldsymbol{v}_3 = \boldsymbol{x}_0 + s\boldsymbol{e}_2. \end{cases}$$

易知 $\boldsymbol{H}(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ 为直角三角形，直角边为 s. 在二维情形下，由(7－2)和(7－3)所构成的单纯形分别如图 7－2，图 7－3 所示.

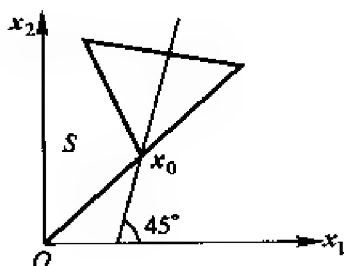


图 7－2

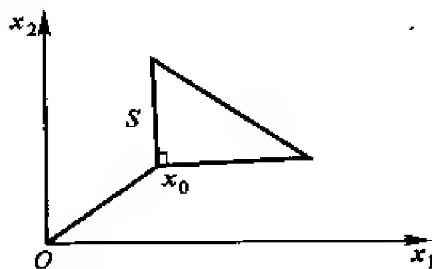


图 7－3

算法 7.1（构造初始单纯形）

① 赋初值 $\boldsymbol{x}_0, s$；

② 计算 p, q：$p = \dfrac{s}{n\sqrt{2}}\left[\sqrt{n+1} + (n-1)\right]$，$q = \dfrac{s}{n\sqrt{2}}\left(\sqrt{n+1} - 1\right)$；

③ 置 $\boldsymbol{v}_1 = \boldsymbol{x}_0$；

④ 对 $i = 2, 3, \cdots, n+1$ 计算　$\boldsymbol{v}_i = \boldsymbol{x}_0 + \boldsymbol{z}_i$；

⑤ 对 $i = 1, 2, \cdots, n+1$ 计算　$f_i = f(\boldsymbol{v}_i)$；

⑥ 对 $i = 1, 2, \cdots, n+1$ 计算　$f_h = \max\limits_{1 \le i \le n+1} \{f_i\}$，$f_l = \min\limits_{1 \le i \le n+1} \{f_i\}$，

$$\boldsymbol{v}_h = \boldsymbol{v}_i (i = h), \ \boldsymbol{v}_l = \boldsymbol{v}_i (i = l);$$

⑦ 计算 $H(v_1,\cdots,v_{h-1},v_{h+1},\cdots,v_{n+1})$ 的中点 $\bar{v}_h = \frac{1}{n}\sum_{\substack{i=1\\i\neq h}}^{n+1} v_i$;

⑧ Stop.

算法 7.1 不仅构造了初始单纯形 $H(v_1,v_2,\cdots,v_{n+1})$,而且分别计算出该单纯形的最好点 v_l(目标函数值最小的顶点)和最坏点 v_h(目标函数值最大的顶点). 还计算了除去最坏点 v_h 后所形成的 $n-1$ 维单纯形 $H(v_1,\cdots,v_{h-1},v_{h+1},\cdots,v_{n+1})$ 的中心 $\bar{v}_h$. 上述工作为实现单纯形替换法作准备.

7.1.2 单纯形替换算法

单纯形替换法的基本思想是,在 $\mathbf{R}^n$ 中选定一个初始点 $v_1=x_0$,按算法 7.1 构造初始单纯形 $H(v_1,v_2,\cdots,v_{n+1})$ 并完成准备工作. 再从初始单纯形出发,采用反射、延伸、压缩、缩小棱长等四个步骤形成新的单纯形以替换原有的单纯形,使新的单纯形不断向极小点靠近,直至搜索到极小点.

算法 7.2(单纯形法)

已知:目标函数 $f(x)$,终止法则的误差限为 ε.

① 准备:执行算法 7.1 得到 $f_h,f_l,v_h,v_l,\bar{v}_h$;

② 反射步(Reflection);

计算最坏点 v_h 的反射点,记为 v_h^R, $v_h^R=\bar{v}_h+\alpha(\bar{v}_h-v_h)$, $\alpha>0$(常取 $\alpha=1$);

计算 v_h^R 对应的目标值 $f_h^R=f(v_h^R)$;

(1) 若 $f_h^R<f_l$,反射最成功,由于 $\bar{v}_h-v_h$ 方向最有利,转入延伸步;

(2) 若在 $i=1,2,\cdots,n+1(i\neq h)$ 中存在某一个 i_0,使 $f_h^R\leqslant f_{i_0}$(除去最坏点外,反射点不比其余所有点都坏). 此时反射成功. 以 v_h^R 取代 v_h 形成新的单纯形

$$H(v_1,\cdots,v_{h-1},v_h^R,v_{h+1},\cdots,v_{n+1});$$

(3) 若 $f_h^R>f_i(i=1,\cdots,n+1,i\neq h)$,但 $f_h^R\leqslant f_h$(除去最坏点外,反射点比其余所有点都坏),反射不大成功,转入压缩步;

(4) 若 $f_h^R>f_i(i=1,\cdots,n+1)$,反射失败,转入压缩步.

③ 延伸步(Expansion)

计算反射点 v_h^R 的延伸点,记为 v_h^E, $v_h^E=\bar{v}_h+\gamma(v_h^R-v_h)$, $\gamma>1$(常取 $\gamma=2$);

计算 $f_h^E=f(v_h^E)$;

(1) 若 $f_h^E < f_l$，延伸成功，以 v_h^E 取代 v_h 形成新的单纯形

$$H(v_1, \cdots, v_{h-1}, v_h^E, v_{h+1}, \cdots, v_{n+1});$$

(2) 若 $f_h^E \geq f_l$，延伸失败，但若 $f_h^R < f_l$，则以 v_h^R 取代 v_h 形成新的单纯形

$$H(v_1, \cdots, v_{h-1}, v_h^R, v_{h+1}, \cdots, v_{n+1});$$

④ 压缩步(Contraction)

(1) 若属于反射步情形(3)，则计算中心 $\bar{v}_h$ 的压缩点，记为 v_h^C，

$$v_h^C = \bar{v}_h + \beta(v_h^R - \bar{v}_h), \beta \in (0,1)\left(常取 \beta = \frac{1}{2}\right);$$

(2) 若属于反射步情形(4)，则舍去反射点 v_h^R，计算中心点 $\bar{v}_h$ 的压缩点 v_h^C，令

$$v_h^C = \bar{v}_h + \beta(v_h - \bar{v}_h), \beta \in (0,1)\left(常取 \beta = \frac{1}{2}\right);$$

并计算 $f_h^C = f(v_h^C)$，

若 $f_h^C \leq \min(f_h, f_h^R)$，则视压缩步成功，以 v_h^C 取代 v_h，形成新的单纯形

$$H(v_1, \cdots, v_{h-1}, v_h^C, v_{h+1}, \cdots, v_{n+1});$$

若 $f_h^C > \min(f_h, f_h^R)$，则转入减小棱长步.

⑤ 减小棱长步

令 $\hat{v}_i = \frac{1}{2}(v_i + v_l)(i = 1, \cdots, n+1)$，形成新的单纯形 $H(\hat{v}_1, \hat{v}_2, \cdots, \hat{v}_{n+1})$. 该步将原单纯形各棱长减小一半，且保留最好点 v_l 为顶点，形成新的单纯形.

在上述各步骤中，当形成新的单纯形后，再重复以上各步，从而形成单纯形序列，使之朝着极小点运动.

算法 7.2 的终止条件：

在执行算法 7.2 中，每当得到一个新的单纯形 $H(v_1, v_2, \cdots, v_{n+1})$，检查如下不等式：

$$\frac{1}{n}\sum_{i=1}^{n+1}(f_i - \bar{f})^2 < \varepsilon ?$$

若满足，则结束迭代；否则继续迭代，

其中 $f_i = f(v_i), \bar{f} = \frac{1}{n+1}\sum_{i=1}^{n+1} f_i$.

在 $\mathbf{R}^2$ 的情形，算法 7.2 中的②—⑤步如图 7－4 所示：

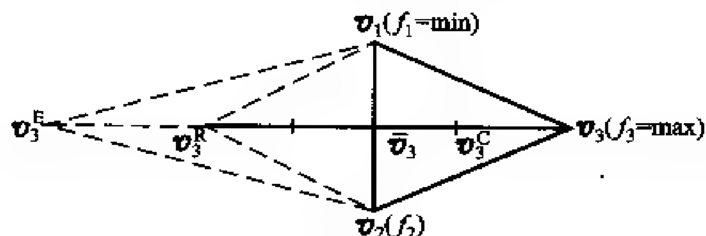


图 7－4

下面给出的三个框图，分别表示算法 7.2 的准备步（图 7－5）；反射、延伸、压缩和减少棱长步（图 7－6）；算法的终止法则（图 7－7）.

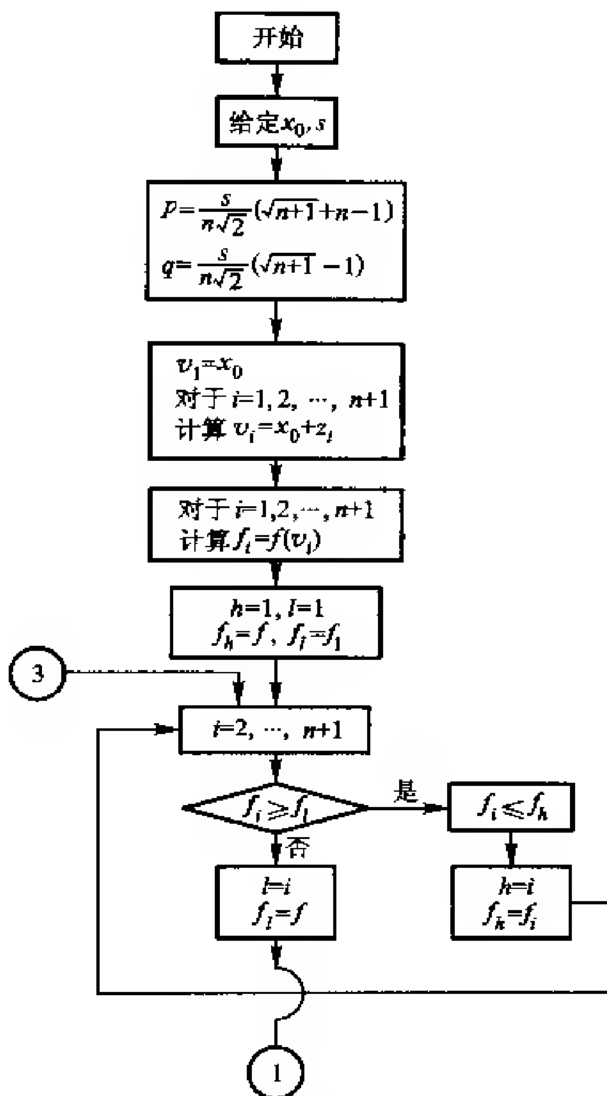


图 7－5　单纯形法的准备步

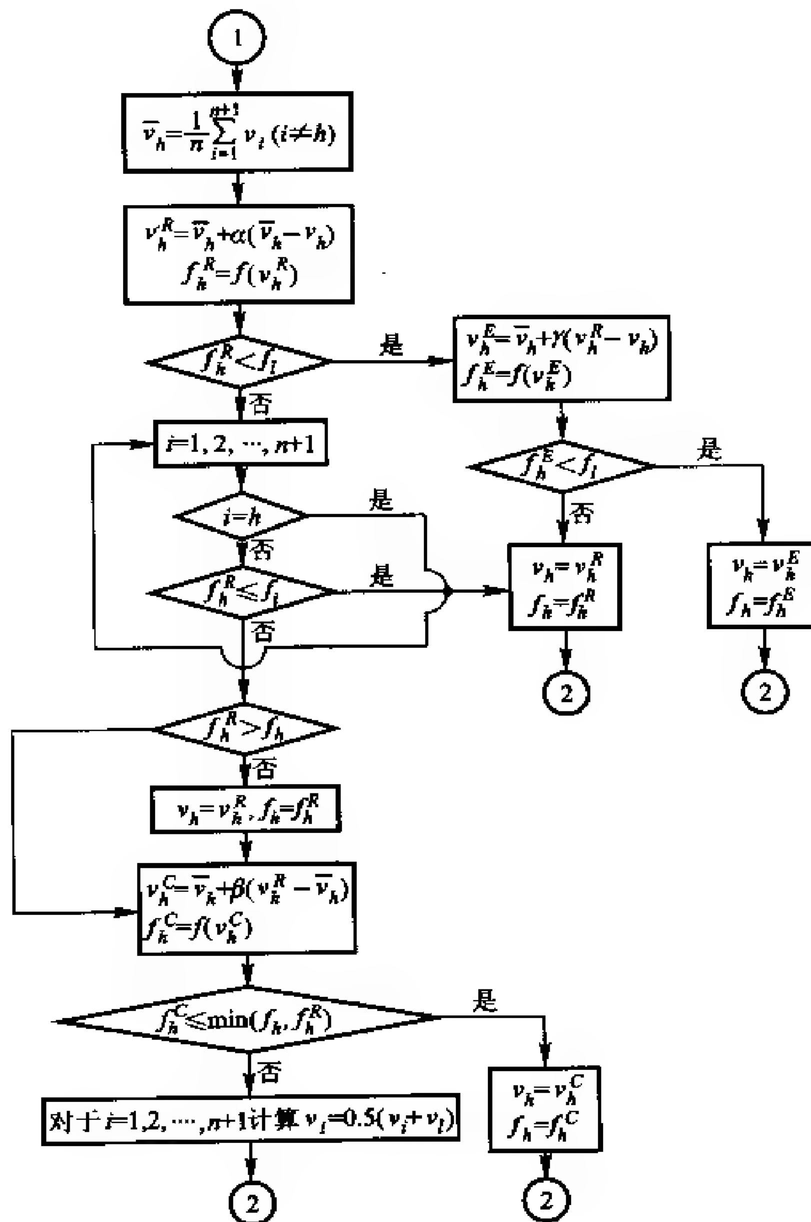


图 7－6 单纯形法的反射、延伸、压缩和减小棱长步

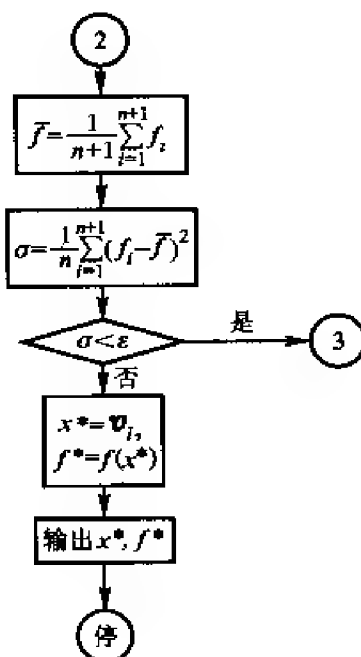


图 7－7　单纯形法的终止法则

7.2　方向加速法

方向加速法是由 Powell(1964 年)提出来的,因此又称 Powell 方法.迄今为止,它是直接方法中最有效的算法.

对于正定二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c, \tag{7-4}$$

在不用导数的前提下,Powell 方法逐次迭代构造 $\boldsymbol{A}$ 共轭方向系,因此 Powell 方法本质上属于共轭方向法,具有二次收敛性.用于非二次函数时一般也具有较快的收敛速度.

7.2.1　基本定理及 Powell 基本算法

定理 7.1　设(1) $f(\boldsymbol{x})$是形如(7－4)的 n 元正定二次函数;

(2) $\boldsymbol{p}_i(i=0,1,\cdots,m-1,m<n)$为 $\boldsymbol{A}$－共轭向量系;

(3) 线性流形 $\boldsymbol{M}$ 定义如下:

$$\boldsymbol{M} = \{\boldsymbol{x}\in\mathbf{R}^n \mid \boldsymbol{x} = \boldsymbol{x}_0 + \sum_{i=0}^{m-1} c_i\boldsymbol{p}_i, \boldsymbol{x}_0\in\mathbf{R}^n\},$$

则 $f(\boldsymbol{x})$在 $\boldsymbol{M}$ 上的极小点 $\boldsymbol{x}_M^*$ 可以沿着每一个方向 $\boldsymbol{p}_i(i=0,1,\cdots,m)$以任意次序进行一维寻查而得到.(证明略)

这个定理的几何意义(三维情形)如图 7－8 所示. 从 $x_0 \in \mathbf{R}^3$ 出发,依次沿 p_0, p_1 作一维寻查得极小点 x_2^*;仍从 x_0 出发,依次沿 p_1, p_0 作一维寻查仍得同一极小点 x_2^*.

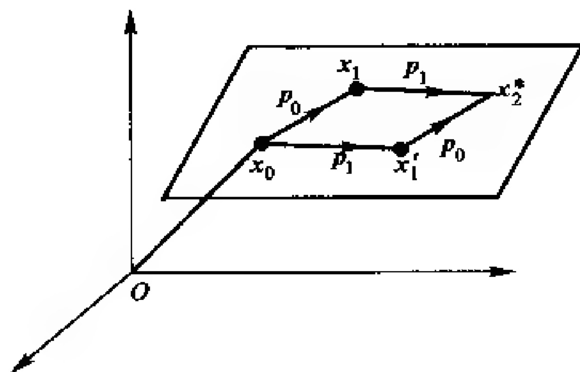


图 7－8

定理 7.2 设

(1) $f(x)$ 为形如(7－4)的正定二次函数;

(2) 任给 $v \in \mathbf{R}^n, x_1, x_2 \in \mathbf{R}^n (x_1 \neq x_2)$;

(3) 分别从 x_1, x_2 出发,沿方向 v 作一维寻查得到极小点 u_1, u_2(如图 7－9 所示).

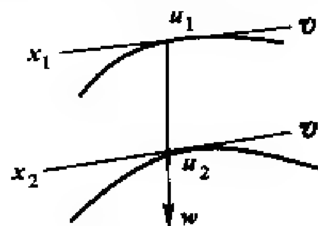


图 7－9

$$\begin{cases} u_1 = x_1 + \alpha_1^* v, \\ \alpha_1^* : \min\limits_{\alpha>0} f(x_1 + \alpha v); \end{cases} \quad \begin{cases} u_2 = x_2 + \alpha_2^* v, \\ \alpha_2^* : \min\limits_{\alpha>0} f(x_2 + \alpha v). \end{cases}$$

令 $w = u_2 - u_1$,则 $w^T A v = 0$,即 w 与 v A－共轭.

证 据一维寻查的性质(见第 3 章(3－2)式)有

$$\nabla f(x_1)^T v = 0, \quad \nabla f(x_2)^T v = 0.$$

以上两式相减得

$$v^T(\nabla f(x_2) - \nabla f(x_1)) = 0. \tag{7－5}$$

由 $\nabla f(x) = Ax + b$ 可得

$$\nabla f(x_2) - \nabla f(x_1) = A(x_2 - x_1) = Aw,$$

代入式(7－5)得

$$v^T A w = 0.$$

上述定理为 Powell 方法构造共轭方向据供了理论依据.

例 7.1 设 $f(x) = f(x_1, x_2, x_3)$ 为正定二次函数,给定三个线性独立的向量:$\xi_1 = (1,0,0)^T, \xi_2 = (0,1,0)^T, \xi_3 = (0,0,1)^T$,任给 $x_0 \in \mathbf{R}^3$,从

x_0 出发，按表 7－2 所示的方式作一维寻查.

表 7－2

	ξ_1	ξ_2	ξ_3	$\xi = x_3 - x_0$
x_0	$x_1 = x_0 + \alpha_1\xi_1$ $\xi_1^{(1)} = \xi_2$	$x_2 = x_1 + \alpha_2\xi_2$ $\xi_2^{(1)} = \xi_3$	$x_3 = x_2 + \alpha_3\xi_3$ $\xi_3^{(1)} = \xi$	$x_4 = x_3 + \alpha\xi$ $\xi^{(1)} = x_3^{(1)} - x_0^{(1)}$
$x_0^{(1)} = x_4$	$x_1^{(1)} = x_0^{(1)}$ $+ \alpha_1^{(1)}\xi_1^{(1)}$ $\xi_1^{(2)} = \xi_2^{(1)}$	$x_2^{(1)} = x_1^{(1)}$ $+ \alpha_2^{(1)}\xi_2^{(1)}$ $\xi_2^{(2)} = \xi_3^{(1)}$	$x_3^{(1)} = x_2^{(1)}$ $+ \alpha_3^{(1)}\xi_3^{(1)}$ $\xi_3^{(2)} = \xi^{(1)}$	$x_4^{(1)} = x_3^{(1)}$ $+ \alpha^{(1)}\xi^{(1)}$ $\xi^{(2)} = x_3^{(2)} - x_0^{(2)}$
$x_0^{(2)} = x_4^{(1)}$	$x_1^{(2)} = x_0^{(2)}$ $+ \alpha_1^{(2)}\xi_1^{(2)}$	$x_2^{(2)} = x_1^{(2)}$ $+ \alpha_2^{(2)}\xi_2^{(2)}$	$x_3^{(2)} = x_2^{(2)}$ $+ \alpha_3^{(2)}\xi_3^{(2)}$	$x_4^{(2)} = x_3^{(2)}$ $+ \alpha^{(2)}\xi^{(2)}$

第一轮从 x_0 出发，依次沿 ξ_1, ξ_2, ξ_3 作一维寻查得 x_3，再沿模式方向 $\xi = x_3 - x_0$ 作一维寻查得 x_4；

第二轮从 $x_0^{(1)} = x_4$ 出发，依次沿 $\xi_1^{(1)}, \xi_2^{(1)}, \xi_3^{(1)}$ 作一维寻查（用 ξ 顶替 ξ_1）得 $x_3^{(1)}$，再沿模式方向 $\xi^{(1)} = x_3^{(1)} - x_0^{(1)}$ 作一维寻查得 $x_4^{(1)}$；

第三轮从 $x_0^{(2)} = x_4^{(1)}$ 出发，依次沿 $\xi_1^{(2)}, \xi_2^{(2)}, \xi_3^{(2)}$ 作一维寻查（用 $\xi^{(1)}$ 顶替 ξ_2）得 $x_3^{(2)}$，再沿模式方向 $\xi^{(2)} = x_3^{(2)} - x_0^{(2)}$ 作一维寻查得 $x_4^{(2)}$.

可以证明正定二次函数 $f(x)$ 的极小点 $x^* = x_4^{(2)}$.

证　首先证明向量组 $\xi_2^{(2)}, \xi_3^{(2)}, \xi^{(2)}$ 即 $\xi, \xi^{(1)}, \xi^{(2)}$ 是 A－共轭方向组. 事实上，据表 7－2 可知，

从 x_3 出发，沿 ξ 作一维寻查得到 x_4，又从 $x_2^{(1)}$ 出发，沿 $\xi_3^{(1)} = \xi$ 作一维寻查得到 $x_3^{(1)}$，据定理 7.2，

$$x_3^{(1)} - x_4 = x_3^{(1)} - x_0^{(1)} = \xi^{(1)} \text{应与 } \xi A - \text{共轭；}$$

从 $x_3^{(1)}$ 出发，沿 $\xi^{(1)}$ 作一维寻查得 $x_4^{(1)}$，又从 $x_2^{(2)}$ 出发沿 $\xi_3^{(2)} = \xi^{(1)}$ 作一维寻查得 $x_3^{(2)}$，据定理 7.2，

$$x_3^{(2)} - x_4^{(1)} = x_3^{(2)} - x_0^{(2)} = \xi^{(2)} \text{应与 } \xi^{(1)} A - \text{共轭；}$$

从 $x_2^{(1)}$ 出发，沿 $\xi_3^{(1)} = \xi$ 作一维寻查得 $x_3^{(1)}$，又从 $x_1^{(2)}$ 出发，沿 $\xi_2^{(2)} = \xi_3^{(1)} = \xi$ 作一维寻查得 $x_2^{(2)}$，而

$$\begin{aligned} x_2^{(2)} - x_3^{(1)} &= (x_3^{(2)} - \alpha_3^{(2)}\xi_3^{(2)}) - (x_4^{(1)} - \alpha^{(1)}\xi^{(1)}) \\ &= (x_3^{(2)} - x_4^{(1)}) + (\alpha^{(1)}\xi^{(1)} - \alpha_3^{(2)}\xi^{(1)}) \\ &= (x_3^{(2)} - x_0^{(2)}) + (\alpha^{(1)} - \alpha_3^{(2)})\xi^{(1)} \end{aligned}$$

$$=\xi^{(2)}+(\alpha^{(1)}-\alpha_3^{(1)})\xi^{(1)}.$$

据定理 7.2，$\xi^{(2)}+(\alpha^{(1)}-\alpha_3^{(1)})\xi^{(1)}$ 应为 ξA－共轭，即

$$(\xi^{(2)}+(\alpha^{(1)}-\alpha_3^{(1)})\xi^{(1)})^{\mathrm{T}}A\xi=0,$$

$$\xi^{(2)}A\xi+(\alpha^{(1)}-\alpha_3^{(1)})\xi^{(1)\mathrm{T}}A\xi=0$$

由 $\xi^{(1)^{\mathrm{T}}}A\xi=0$ 得 $\xi^{(2)^{\mathrm{T}}}A\xi=0$，即 $\xi^{(2)}$ 与 ξA－共轭.

综上，$\xi,\xi^{(1)},\xi^{(2)}$ 互为 A－共轭，亦即向量组 $\xi_2^{(2)},\xi_3^{(2)},\xi^{(2)}$ 为 A－共轭组. 据第 5 章定理 5.4，在表 7－2 第三轮维寻中，从 $x_1^{(2)}$ 出发，依次沿 $\xi_2^{(2)},\xi_3^{(2)},\xi^{(2)}$ 作一维寻查所得 $x_4^{(2)}$ 应为正定二次函数 $f(x)$ 的极小点，即 $x^*=x_4$.

将例 7.1 的计算步骤和结论推广到 n 元正定二次函数便得到 Powell 基本算法.

算法 7.3（正定二次函数的 Powell 基本算法）

设 $f(x)$ 为正定二次函数，给定初始点 x_0

① 置 $k=0$；

② 令 $\xi_i=e_i(i=1,\cdots,n)$；

③ 对 $i=1,\cdots,n$ 分别计算

$$\begin{cases}x_i=x_{i-1}+\alpha_i\xi_i,\\ \alpha_i:\min\limits_{\alpha>0}f(x_{i-1}+\alpha\xi_i),\end{cases}$$

④ 令 $\xi=x_n-x_0$，计算

$$\begin{cases}x_{n+1}=x_n+\alpha_{n+1}\xi,\\ \alpha_{n+1}:\min f(x_n+\alpha\xi);\end{cases}$$

⑤ 检查 $k=n-1$？

是，$x_{n+1}\Rightarrow x^*$，Stop，

否，转⑥

⑥ $\xi_{i+1}\Rightarrow\xi_i(i=1,2,\cdots,n-1)$，

$\xi\Rightarrow\xi_n$，

$x_{n+1}\Rightarrow x_0$，

$k+1\Rightarrow k$，转③.

算法 7.3 如表 7－3 所示.

表 7-3

		$\xi_1=e_1$	$\xi_2=e_2$	…	$\xi_n=e_n$	$\xi=x_n-x_0$
$k=0$	x_0	$x_1=x_0+\alpha_0\xi_1$	$x_2=x_1+\alpha_1\xi_2$	…	$x_n=x_{n-1}+\alpha_n\xi_n$	$x_{n+1}=x_n+\alpha_{n+1}\xi$
		$\xi_1\Leftarrow\xi_2$	$\xi_2\Leftarrow\xi_3$	…	$\xi_n\Leftarrow\xi$	$\xi=x_n-x_0$
$k=1$	$x_0=x_{n+1}$	x_1	x_2	…	x_n	x_{n+1}
	…	…	…	…	…	…
		$\xi_1\Leftarrow\xi_2$	$\xi_2\Leftarrow\xi_3$	…	$\xi_n\Leftarrow\xi$	$\xi=x_n-x_0$
$k=n-1$	$x_0=x_{n+1}$	x_1	x_2	…	x_n	$x_{n+1}=x^*$

算法 7.3 的特点是，每一轮寻查后，寻查方向时总是用模式方向（加速方向）ξ 替换 ξ_1，形成新的方向组 $\xi,\xi_2,\cdots,\xi_n$. 但据定理 7.1，维查方向的次序可以任意交换，故采用依次替换的方式形成新的寻查方向组：$\xi_1\Leftarrow\xi_2,\xi_2\Leftarrow\xi_3,\cdots,\xi_n\Leftarrow\xi$.

从理论上讲，通过 n 轮寻查所得到的新的方向组应为 A－共轭方向系. 但是在实际中会出现异常现象，即所形成的新的寻查方向组非 A－共轭方向系. 甚至是线性相关的，即使对正定二次函数也会出现这种情况.

例 7.2 设 $f(x)=f(x_1,x_2)$，初始点 x_0，极小点 x^*.

	$\xi_1=e_1$	$\xi_2=e_2$	$\xi=x_2-x_0$
x_0	$x_1=x_0$	$x_2=x_1+\alpha_2\xi_2$	$x_3=x_2+\alpha\xi$

如图 7-10 所示，由于第一轮寻查产生退化 $x_1=x_0$，致使 ξ_2,ξ 成为同一方向，即第二轮寻查方向 $\xi_1\Leftarrow\xi_2,\xi_2\Leftarrow\xi$ 仍是同一方向，即是说搜索点只是在纵轴上移动，永远无法达到极小点 x^*.

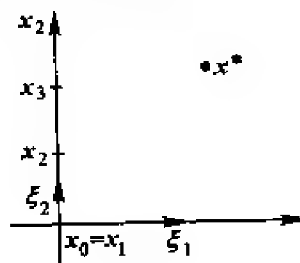


图 7-10

7.2.2 Powell 算法的方向调整原理

在 Powell 基本算法 7.3 中，方向调整的原则是：获得模式方向 $\xi = x_n - x_0$ 后，下一轮寻查一定要调整方向，且总是用 ξ 去替换 ξ_1，这就是可能产生向量组线性相关的原因所在. 因此，在改进算法中首先给出方向调整的判别准则，当每一轮获得模式方向 $\xi = x_n - x_0$ 后，下一轮是否调整方向？ξ 替换谁？均由判别准则来决定，从而保证每一轮寻查方向组成为"最接近 A - 共轭"的 n 个方向.

引理 7.1 设 A 为 $n \times n$ 阶矩阵，令 $A = (a_1, a_2, \cdots, a_n)$，其中 $a_i = (a_{1i}, a_{2i}, \cdots, a_{ni})^T (i = 1, 2, \cdots, n)$，则 $|\det(A)| \leqslant \prod_{i=1}^{n} \| a_i \|_2$.

证 $A^T = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix}$，其中 $a_i^T = (a_{1i}, a_{2i}, \cdots, a_{ni})$，

据 Hadamard 不等式，$|\det(A^T)| \leqslant \prod_{i=1}^{n} \| a_i^T \|_2$，

又因为 $\det(A) = \det(A^T)$，且 $\| a_i^T \|_2 = \| a_i \|_2$，

所以 $|\det(A)| \leqslant \prod_{i=1}^{n} \| a_i \|_2$.

定理 7.3 设(1) A 是 $n \times n$ 对称正定矩阵；

(2) $\xi_i (i = 1, \cdots, n)$ 线性无关，且满足 $\xi_i^T A \xi_i = 1$（称 $\xi_1, \cdots, \xi_n$ 为 A 规格化向量组）；

(3) 令 $X = (\xi_1, \cdots, \xi_n)$，

则 $|\det(X)|$ 取最大值的充要条件是 $\xi_i (i = 1, \cdots, n)$ 为 A - 共轭系，即

$$\xi_i^T A \xi_j = 0 \quad (i, j = 1, \cdots, n), \quad i \neq j.$$

证 设 $\eta_i (i = 1, \cdots, n)$ 为 A - 共轭且 A 规格化向量组，即

$$(\eta_i, \eta_i)_A = 1, \quad \eta_i^T A \eta_j = 0 \ (i \neq j).$$

令
$$\xi_i = \sum_{j=1}^{n} u_{ij} \eta_j, \tag{7-6}$$

分别令 $U = (u_1, \cdots, u_n)$，其中 $u_i = (u_{i1}, u_{i2}, \cdots, u_{in})^T$，

$$V = (\eta_1, \cdots, \eta_n),$$

据(7-6)，$\xi_i = V u_i (i = 1, \cdots, n)$， (7-7)

所以
$$X = V \cdot U,$$

$$\det(X) = \det(V)\cdot\det(U). \tag{7-8}$$

下证$|\det(U)|\leqslant 1$.

据(7-7),$\xi_i = Vu_i$,$\xi_i^T A\xi_j = u_i^T(V^T AV)u_i$,

而 $V^T AV = \begin{bmatrix}\eta_1^T\\ \vdots \\ \eta_n^T\end{bmatrix}(A\eta_1,\cdots,A\eta_n) = I_n$,

所以 $u_i^T u_i = \xi_i^T A\xi_j = 1$,即$\|u_i\|_2 = 1(i=1,\cdots,n)$.

据引理7.1得

$$|\det(U)| \leqslant \prod_{i=1}^{n}\|u_i\|_2 = 1. \tag{7-9}$$

再由式(7-8)立即得 $|\det(X)| \leqslant |\det(V)|$. (7-10)

综合(7-8),(7-9),(7-10)三式可知

$|\det(X)|$取得最大值$|\det(V)|$的充要条件是$|\det(U)|=1$,即U为正交阵($V^T U = I_n$);而$X^T AX = (VU)^T A(VU) = V^T(V^T AV)U = U^T U$,因此$|\det(X)|$取得最大值的充要条件是$X^T AX = I_n$,即 $\xi_i^T A\xi_j = \begin{cases}1, i=j,\\ 0, i\neq j,\end{cases}$ 亦即$\xi_i(i=1,\cdots,n)$为A规格且A-共轭向量组.

据上述定理,得到搜索方向替换的判别准则:

设某一轮搜索方向为A规格化向量组

$$\xi_1,\cdots,\xi_r,\cdots,\xi_n (1\leqslant r\leqslant n),$$

令

$$X = (\xi_1,\cdots,\xi_r,\cdots,\xi_n),$$

$$X_r = (\xi_1,\cdots,\xi_{r-1},\xi,\xi_{r+1},\cdots,\xi_n),$$

X_m满足$|\det(X_m)| = \max\limits_{1\leqslant r\leqslant n}|\det(X_r)|$,若$|\det(X_m)| > \det(X)$,则原方向组$\xi_1,\cdots,\xi_{m-1},\xi_m,\xi_{m+1},\cdots,\xi_n$应替换为$\xi_1,\cdots,\xi_{m-1},\xi,\xi_{m+1},\cdots,\xi_n$.

此时新的方向组不仅比原方向组而且比所有替换后的方向组都更接近A-共轭.

该方向调整的判别准则同时解决了该不该替换,该如何替换这样两个问题.但是应用上述判别准则的困难在于确定m.因为按$|\det(X_m)| = \max\limits_{1\leqslant r\leqslant n}|\det(X_r)|$确定$m$,其计算量是十分大的,下面的定理给出确定$m$的简便方法.

定理7.4 设(1) $f(x)$是正定二次函数;

(2) $\xi_i(i=1,\cdots,n)$是Powell基本算法中的搜索方向且满足A规格化;

(3) 对于 $i=1,\cdots,n$

$$\begin{cases} x_i = x_{i-1} + \alpha_i \xi_i, \\ \alpha_i : \min\limits_{\alpha>0} f(x_{i-1} + \alpha \xi_i); \end{cases}$$

(4) $\xi = \frac{1}{\mu}(x_n - x_0)$ 且 $\xi^T A \xi = 1$ $\left(\frac{1}{\mu}\text{为实数,称为规格化因子}\right)$;

(5) $X=(\xi_1,\cdots,\xi_r,\cdots,\xi_n)$, $X_r=(\xi_1,\cdots,\xi_{r-1},\xi,\xi_{r+1},\cdots,\xi_n)$,

则有(1) $\det(X_r) = \frac{\alpha_r}{\mu}\det(X)$;

(2) 在下列三组数中 $\{|\alpha_r|\}_{r=1,\cdots,n}$; $\{f(x_{r-1}) - f(x_r)\}_{r=1,\cdots,n}$; $\{|\det(X_r)|\}_{r=1,\cdots,n}$ 各组的最大值应具有同一下标.

据定理 7.4 的结论(2),如果

$$f(x_{m-1}) - f(x_m) = \max_{1\leqslant r\leqslant n}\{f(x_{r-1}) - f(x_r)\}, \tag{7-11}$$

则

$$|\det(X_m)| = \max_{1\leqslant r\leqslant n}|\det(X_r)|, \tag{7-12}$$

反之亦然.

显然,用(7-11)替换(7-12)更容易确定 m.

下面我们给出定理 7.4 证明的要点,请读者自行补充.

证 (1) 由假设(4)

$$\xi = \frac{1}{\mu}(x_n - x_0) = \frac{1}{\mu}\sum_{i=1}^{n}(x_i - x_{i-1}) = \frac{1}{\mu}\sum_{i=1}^{n}\alpha_i\xi_i = \frac{\alpha_i}{\mu}\sum_{i=1}^{n}\xi_i,$$

容易导出 $\det(X_r) = \frac{\alpha_r}{\mu}\det(X)$,即定理 7.4 的结论(1);

(2) 首先证 $\{|\det(X_r)|\}_{r=1,\cdots,n}$ 与 $\{|\alpha_r|\}_{r=1,\cdots,n}$ 的最大值具有同一下标,即 $|\alpha_m| = \max\limits_{1\leqslant r\leqslant n}|\alpha_r|$ 与 $|\det(X_m)| = \max\limits_{1\leqslant r\leqslant n}|\det(X_r)|$ 同时成立.

再证 $\{|\alpha_r|\}_{r=1,\cdots,n}$ 与 $\{f(x_{r-1}) - f(x_r)\}_{r=1,\cdots,n}$ 的最大值具有同一下标,即 $|\alpha_m| = \max\limits_{1\leqslant r\leqslant n}|\alpha_r|$ 与 $\{f(x_{r-1}) - f(x_r)\}$ 同时成立.

综合可得定理 7.4 的结论(2).

定理 7.4 给出了实用的方向调整的判别条件.

7.2.3 Powell 算法方向调整的判别准则

1. 正定二次函数的情形

设 $f(x)$ 为正定二次函数,某一轮搜索方向 $\xi_i(i=1,\cdots,n)$ 为线性独立且是 A 规格化的向量组. 从任给点 x_0 出发,依次沿 $\xi_i(i=1,\cdots,n)$ 作一维寻查,如表 7-4 所示:

表 7-4

	ξ_1	ξ_2	…	ξ_n	$\xi=\frac{1}{\mu}(x_n-x_0)$
x_0	$x_1=x_0+\alpha_1\xi_1$	$x_2=x_1+\alpha_2\xi_2$	…	$x_n=x_{n-1}+\alpha_{n-1}\xi_n$	$x_{n+1}=x_n+\alpha_n\xi$
	$f_1=f(x_1)$	$f_2=f(x_2)$	…	$f_n=f(x_n)$	

其中 $\mu^2=(x_n-x_0)^{\mathrm T}A(x_n-x_0)$，$\alpha_i=-\dfrac{g_i^{\mathrm T}\xi_i}{\xi_i^{\mathrm T}A\xi_i}$.

计算 $f(x_{r-1})-f(x_r)$，$r=1,\cdots,n$.

设 $f(x_{m-1})-f(x_m)=\max\limits_{1\leqslant r\leqslant n}\{f(x_{r-1})-f(x_r)\}$，

若 $\left|\dfrac{\alpha_m}{\mu}\right|\leqslant 1$，据定理 7.4 的结论(1)，有

$$|\det(X_m)|\leqslant|\det(X)|,$$

又据定理 7.4 的结论(2)，有

$$|\det(X)|\geqslant|\det(X_m)|=\max_{1\leqslant r\leqslant n}|\det(X_r)|.$$

故以 ξ 替换 ξ_m 后，$\{|\det(X_r)|\}_{r=1,\cdots,n}$ 均不会增大，因此 ξ 不应该换入，应保持原方向组不变；

若 $\left|\dfrac{\alpha_m}{\mu}\right|>1$，因为 $|\det(X_m)|>|\det(X)|$，故应以 ξ 替换 ξ_m，所形成的新的搜索方向组 $\xi_1,\cdots,\xi_{m-1},\xi_m,\xi_{m+1},\cdots,\xi_n$ 较原方向组更接近 A-共轭.

算法 7.4（正定二次函数的 Powell 改进算法）

① 给定初始点 x_0 及初始函数值 $f_0=f(x_0)$，$\varepsilon>0$；

② 置 $k=0$；

③ 给定初始搜索方向 $\xi_i=e_i(i=1,\cdots,n)$；

④ 对 $i=1,\cdots,n$ 计算

$$\begin{cases} x_i=x_{i-1}+\alpha_i\xi_i, \\ \alpha_i:\ \min\limits_{\alpha>0}f(x_{i-1}+\alpha\xi_i); \end{cases}$$

⑤ $f_i=f(x_i)$；

⑥ 计算规格化因子 μ，$\mu=\sqrt{(x_n-x_0)^{\mathrm T}A(x_n-x_0)}$；

⑦ 令 $\xi_n=\dfrac{1}{\mu}(x_n-x_0)$ 计算

$$\begin{cases} x_{n+1}=x_n+\alpha_{n+1}\xi, \\ \alpha_{n+1}:\ \min f(x_n+\alpha\xi); \end{cases}$$

⑧ 检查 $k=n-1$?

是，$\boldsymbol{x}_{n+1}\Rightarrow\boldsymbol{x}^*$，Stop，

否，转⑨；

⑨ 确定 m 使得

$$f(\boldsymbol{x}_{m-1})-f(\boldsymbol{x}_m)=\max_{1\leqslant r\leqslant n}\{f(\boldsymbol{x}_{r-1})-f(\boldsymbol{x}_r)\};$$

⑩ 检查 $\left|\frac{\alpha_m}{\mu}\right|\leqslant 1$?

是，$\boldsymbol{\xi}_i\Rightarrow\boldsymbol{\xi}_i(i=1,2,\cdots,n)$，

$\boldsymbol{x}_{n+1}\Rightarrow\boldsymbol{x}_0$，

$k+1\Rightarrow k$，转③.

否，对 $i=m,m+1,\cdots,n-1$.

$\boldsymbol{\xi}_{i+1}\Rightarrow\boldsymbol{\xi}_i$，

置$\boldsymbol{\xi}_n=\boldsymbol{\xi}$，

$\boldsymbol{x}_{n+1}\Rightarrow\boldsymbol{x}_0$，

$k+1\Rightarrow k$，转④.

2. 非二次函数的情形

对于非二次函数 $f(\boldsymbol{x})$，计算规格化因子 $\mu(\mu^2=(\boldsymbol{x}_m-\boldsymbol{x}_0)\boldsymbol{A}(\boldsymbol{x}_m-\boldsymbol{x}_0))$时通常用$\nabla^2 f(\boldsymbol{x})$去近似代替矩阵 $\boldsymbol{A}$. 但我们知道计算二阶偏导数矩阵$\nabla^2 f(\boldsymbol{x})$违背了直接搜索方法的宗旨，并且当 $\boldsymbol{x}$ 远离极小点 $\boldsymbol{x}^*$ 处，$\nabla^2 f(\boldsymbol{x})$近似 $\boldsymbol{A}$ 的误差较大. 因此对于非二次函数，方向调整的判别条件不宜使用原有的不等式 $\left|\frac{\alpha_m}{\mu}\right|\leqslant 1$，而需要另辟途径.

现在我们考虑模式方向 $\boldsymbol{\xi}=\boldsymbol{x}_n-\boldsymbol{x}_0$ 上的三个等距的点 $\boldsymbol{x}_0,\boldsymbol{x}_n,2\boldsymbol{x}_n-\boldsymbol{x}_0$，其对应的函数值分别为 $f_1=f(\boldsymbol{x}_0)$，$f_2=f(\boldsymbol{x}_n)$，$f_3=f(2\boldsymbol{x}_n-\boldsymbol{x}_0)$ 过 $(\boldsymbol{x}_0,f_1)$，$(\boldsymbol{x}_n,f_2)$，$(2\boldsymbol{x}_n-\boldsymbol{x}_0,f_3)$ 三点作抛物线，设其极小点为 $\hat{\boldsymbol{x}}$，极小值为 $\hat{f}$. 于是有

$$\hat{\boldsymbol{x}}=\boldsymbol{x}_n-\frac{f_3-f_1}{2(f_1-2f_2+f_3)}\boldsymbol{\xi},$$

$$\hat{f}=f_2-\frac{(f_1-f_3)^2}{8(f_1-2f_2+f_3)}.$$

讨论以下两种情形：

情形 1：$f_1-2f_2+f_3<0$.

当 $f(\boldsymbol{x})$ 为正定二次函数时，由于 $f(\boldsymbol{x})$ 有极小点，不会出现情形 1，但当 $f(\boldsymbol{x})$ 非二次函数时，情形 1 是可能出现的. 此时 $\boldsymbol{\xi}$ 方向上将出现 $f_3<f_2<f_1$.

事实上，$f_2<f_1$ 是显然的，而由两个不等式 $f_1-2f_2+f_3<0$，$f_2<f_1$ 将导出 $f_3<f_2$.

可见 ξ 是一个较好的下降方向，因此可用 ξ 来替换 $\xi_i(i=1,\cdots,n)$ 中的 ξ_m.

情形 2：$f_1-2f_2+f_3>0$，当 $f_3\geqslant f_1$ 时，取 $\mu=\sqrt{2(f_1-\hat{f})}+\sqrt{2(f_2-\hat{f})}$，可以推证 $|\alpha_i|\leqslant|\mu|(i=1,\cdots,n)$，即 $\left|\dfrac{\alpha_i}{\mu}\right|\leqslant 1(i=1,\cdots,n)$，从而搜索方向 $\xi_1,\cdots,\xi_n$ 应保持不变，

当 $f_3<f_1$ 时，取 $\mu=\sqrt{2(f_1-\hat{f})}+\sqrt{2(f_2-\hat{f})}$，

令 $\Delta=\max\limits_{1\leqslant i\leqslant n}\{f(x_{i-1})-f(x_i)\}=f(x_{m-1})-f(x_m)$，

$$\frac{\alpha_m}{\mu}=\frac{\sqrt{\Delta}}{\sqrt{2(f_1-\hat{f})}-\sqrt{2(f_2-\hat{f})}},$$

因此，当 $f_3<f_1$ 且 $\dfrac{\sqrt{\Delta}}{\sqrt{2(f_1-\hat{f})}-\sqrt{2(f_2-\hat{f})}}<1$ 时，原方向组 $\xi_i(i=1,\cdots,n)$ 应保持不变；而当 $f_3<f_1$ 且 $\dfrac{\sqrt{\Delta}}{\sqrt{2(f_1-\hat{f})}-\sqrt{2(f_2-\hat{f})}}\geqslant 1$ 时，则应以 ξ 替换 ξ_m.

综上所述，$\xi=\dfrac{1}{\mu}(x_n-x_0)$ 是否替换 ξ_m，可细分为以下四种情形讨论：

(1) 若 $f_1-2f_2+f_3<0$（显然 $f_3<f_1$），则用 ξ 替换 ξ_m；

(2) 若 $f_1-2f_2+f_3>0$ 且 $f_3\geqslant f_1$，则原方向组 $\xi_i(i=1,\cdots,n)$ 应保持不变；

(3) 若 $f_1-2f_2+f_3>0$，$f_3<f_1$ 且 $\sqrt{\Delta}\geqslant\sqrt{2(f_1-\hat{f})}-\sqrt{2(f_2-\hat{f})}$，则用 ξ 替换 ξ_m；

(4) 若 $f_1-2f_2+f_3>0$，$f_3<f_1$ 且 $\sqrt{\Delta}<\sqrt{2(f_2-\hat{f})}-\sqrt{2(f_1-\hat{f})}$，则原方向组 $\xi_i(i=1,\cdots,n)$ 应保持不变.

上述四种情形如图 7-11 所示：

还可以证明，利用判别式

$$(f_1-2f_2+f_3)(f_1-f_2-\Delta)^2\geqslant\frac{\Delta}{2}(f_1-f_3)^2$$

可以将图 7－11 简化为图 7－12.

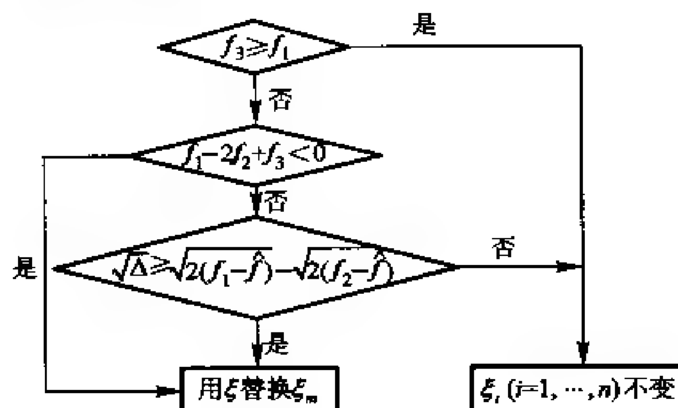


图 7－11

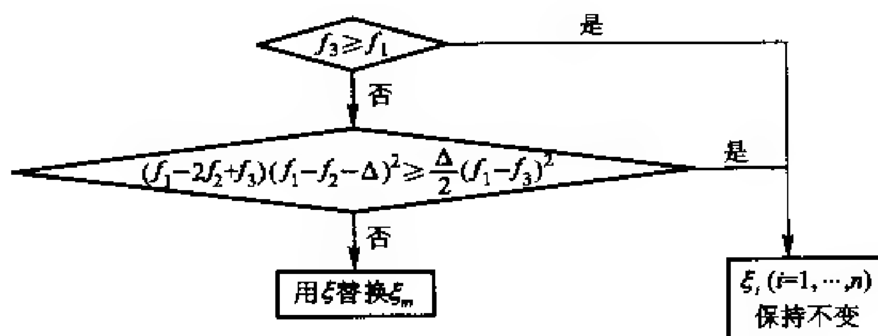


图 7－12

算法 7.5（一般函数的改进算法）

① 给定初始点 x_0，收敛精度 $\varepsilon>0$；

② 给定初始搜索方向 $\xi_i=e_i(i=1,\cdots,n)$；

③ 给定初始规格化因子 $s=\bar{s}=1$；

④ 计算 $f_1=f(x_0)$；

⑤ 对 $i=1,\cdots,n$ 计算

$$\begin{cases} x_i = x_{i-1}+\alpha_i\ p_i, \\ p_i = s\xi_i, \\ \alpha_i:\ \min f(x_{i-1}+\alpha\xi_i); \end{cases}$$

⑥ 计算 $f_2=f(x_n)$；

⑦ 确定 m 使得 $f(x_{m-1})-f(x_m)=\max\limits_{1\leqslant i\leqslant n}|f(x_{i-1})-f(x_i)|$，

$$令\ \Delta=f(x_{m-1})-f(x_m)；$$

⑧ 计算 $f_3=f(2x_n-x_0)$；

⑨ 检查 $f_3<f_1$？

否，$\boldsymbol{p}_i \Rightarrow \boldsymbol{p}_i (i=1,\cdots,n)$转⑬，

是，转⑩；

⑩ 检查$(f_1-2f_2+f_3)(f_1-f_2-\Delta)^2 \geqslant \frac{\Delta}{2}(f_1-f_3)^2$？

是，$\boldsymbol{p}_i \Rightarrow \boldsymbol{p}_i (i=1,\cdots,n)$转⑬，

否，转⑪；

⑪ 作搜索方向替换(去掉$\boldsymbol{p}_m$，换入模式方向$\frac{s}{\|\boldsymbol{x}_n-\boldsymbol{x}_0\|_2}(\boldsymbol{x}_n-\boldsymbol{x}_0)$)；

即对$i=m,m+1,\cdots,n-1$，$\boldsymbol{p}_{i+1} \Rightarrow \boldsymbol{p}_i$，

置$\boldsymbol{p}_n = \frac{s}{\|\boldsymbol{x}_n-\boldsymbol{x}_0\|_2}(\boldsymbol{x}_n-\boldsymbol{x}_0)$；

⑫ 计算

$$\begin{cases} \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha^* \boldsymbol{p}_n, \\ \alpha^* : \min\limits_{\alpha>0} f(\boldsymbol{x}_n + \alpha \boldsymbol{p}_n), \end{cases}$$

$$f_2 = f(\boldsymbol{x}_{n+1});$$

⑬ 检查H－终止条件？

若满足，则输出$\boldsymbol{x}_{n+1}$，f_2，

否则转⑭；

⑭ 修改规格化因子

$$s = \min\{\bar{s}, 0.4 \cdot |f_1 - f_2|^{\frac{1}{2}}\};$$

⑮ 重赋初值$\boldsymbol{x}_{n+1} \Rightarrow \boldsymbol{x}_0$，$f_2 \Rightarrow f_1$ 转⑤.

第 8 章　线性最小二乘法

8.1　观测数据的最小二乘拟合

引例

设通过实验观测手段获得 5 组数据(表 8－1)：

表 8－1

t_i	2	4	5	8	9
y_i	2.01	2.98	3.50	5.02	5.47

它反映出变量 t,y 之间的某种数值对应关系. 现确定数据所表示的函数 $y=y(t)$ 的一种近似表达式.

首先描出观测数据的散点图(图 8－1),上述 5 点在平面上近似地分布在一条直线上,可以假定近似表达式为线性函数

$$y^*(t) = a + bt,$$

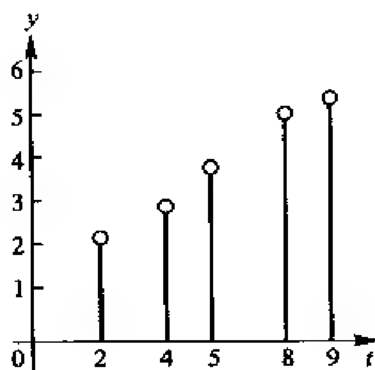


图 8－1

其中参数 a,b 应满足如下关系式

$$\begin{cases} a+2b=2.01, \\ a+4b=2.98, \\ a+5b=3.50, \\ a+8b=5.02, \\ a+9b=5.47. \end{cases} \tag{8-1}$$

显然式(8 - 1)为一超定方程组，由于 5 点并非真正共线，因而方程组(8 - 1)是不相容的，即是说不能由式(8 - 1)直接解得 a,b 的值.

可用如下一些粗略方法确定 a,b 的近似值.

1. 选点法：即从式(8 - 1)中任取两个方程求 a,b. 例如取

$$\begin{cases} a + 2b = 2.01, \\ a + 4b = 2.98, \end{cases}$$

求解得　$a = 1.0400, b = 0.4850$.

取

$$\begin{cases} a + 4b = 2.98, \\ a + 5b = 3.50, \end{cases}$$

得解　$a = 0.9003,\ b = 0.5200$

2. 平均法：为了充分利用已给的数据，将式(8 - 1)中的 5 个方程分成 2 组. 取平均最后得 2 个方程，联立求解 a,b.

例如，将式(8 - 1)中的前三个方程划为一组，余下的为一组，取平均后得 2 个方程：

$$\begin{cases} a + 3.67b = 2.83, \\ a + 8.5b = 5.245, \end{cases}$$

求解得　$a = 0.995,\ b = 0.5$.

平均法较选点法效果会好一些，但是存在如下两个问题：

第一，方法本身没有给出评价解的好坏的标准，什么样的解才算“最好”？

第二，方法也未指出用何种办法求得“最好”的解？

为了解决上述问题，我们引入与观测数据有关的一个重要概念.

8.1.1　残差

对于引例中所给数据

$$(t_i, y_i),\ i = 1,2,\cdots,5,$$

令 y_i 表示观测值，$y_i^* = a^* + b^* t_i$ 表示估计值. 由于 y_i 有误差，$y_i^* = a^* + b^* t_i$ 又是近似表达式，故一般地

$$y_i^* \neq y_i,\ i = 1,2,\cdots,5.$$

记　$$\gamma_i = y_i^* - y_i = (a^* + b^* t_i) - y_i,\ i = 1,2,\cdots,5, \quad (8-2)$$

称 γ_i 为观测数据 y_i 的残差. 由于 t_i, y_i 为已知，因此，γ_i 为参数 a^*, b^* 的函数，即

$$\gamma_i = \gamma_i(a^*, b^*),\ i = 1,2,\cdots,5, \quad (8-3)$$

由此得到以 γ_i 为分量的残差向量

$$\gamma(a^*,b^*) = (\gamma_1,\gamma_2,\cdots,\gamma_5)^T, \tag{8-4}$$

并且,用 γ 的长 $\|\gamma\|$ 作为评价参数 a^*,b^* 好坏的标准.因此我们避开由式(8-1)确定 a^*,b^* 的值,转而求如下无约束极小问题

$$\min\|\gamma(a,b)\|, \tag{8-5}$$

即对 $\forall (a,b)^T \in \mathbf{R}^2$,确定 $(a^*,b^*)^T$,使得

$$\min\|\gamma(a,b)\| = \|\gamma(a^*,b^*)\|,$$

极小问题式(8-5)的解 $(a^*,b^*)^T$ 是在任意范数 $\|\cdot\|$ 意义下的最好解.

为了避免绝对值运算,通常采用欧氏范数,即求解如下无约束极小问题

$$\min\|\gamma(a,b)\|_2^2,$$

其中

$$\|\gamma(a,b)\|_2^2 = \sum_{i=1}^{5}\gamma_i^2(a,b). \tag{8-6}$$

令残量平方和达到最小来确定 a^*,b^*,即所谓观测数据的最小二乘拟合,式(8-6)的解称为超定方程组式(8-1)的最小二乘解.

8.1.2 最小二乘拟合的数学模型

设给出 m 组数据

$$(t_i,y_i),\ i = 1,2,\cdots,m, \tag{8-7}$$

其中 $t_i \neq t_j (i\neq j)$, $y(t_i) = y_i$.

给定一个有限的函数系

$$\{\phi_j(t)\},\ j = 1,2,\cdots,n\ (m > n), \tag{8-8}$$

且假定对于 $t_i \in [a,b]$, $\{\phi_j(t_i)\} (j=1,2,\cdots,n)$ 都是线性独立的.我们的目的是根据已知数据(8-7)用函数系式(8-8)的线性组合在某种意义下来近似表示函数 $y(t)$.

即令

$$y(t,x_1^*,\cdots,x_n^*) = \sum_{j=1}^{n} x_j^*\phi_j(t), \tag{8-9}$$

或简记为 $y(t,\boldsymbol{x}^*) = \sum_{j=1}^{n} x_j^*\phi_j(t)$, $\boldsymbol{x}^* = (x_1^*,\cdots,x_n^*)^T$.

常用的函数系(8-8)有幂函数系 $\{t^j\}$,三角函数系 $\{\sin(jt)\}$,指数函数系 $\{e^{\lambda_j t}\}$ 等.

表达式(8-9)对参数 $x_j^* (j=1,2,\cdots,n)$ 而言是线性的,这是一种线性数学模型.

引入数据式(8-7)的残差向量 $\gamma(\boldsymbol{x}^*)$

$$\gamma(\boldsymbol{x}^*) = (\gamma_1,\gamma_2,\cdots,\gamma_m)^T,$$

其分量(残差)

$$\gamma_i = \gamma_i(\boldsymbol{x}^*) = y(t_i,\boldsymbol{x}^*) - y_i$$

$$= \sum_{j=1}^{n} x_j^* \phi_j(t_i) - y_i, \ i = 1,2,\cdots,m. \qquad (8-10)$$

我们的目的是，按照最小二乘准则来确定式(8－10)中的参数 x_j^* $(j=1,2,\cdots,n)$，即是求解如下无约束极小问题

$$\min_{x\in\mathbf{R}^n} \| \gamma(x) \|_2^2, \qquad (8-11)$$

其中 $\min_{x\in\mathbf{R}^n} \| \gamma(x) \|_2^2 = \sum_{i=1}^{m} \gamma_i^2(x) = \sum_{i=1}^{m} \left[\sum_{j=1}^{n} x_j \phi_j(t_i) - y_i \right]^2,$

称式(8－11)的解 $x^* = (x_1^*,\cdots,x_n^*)^T$ 叫最小二乘解.

通过求最小二乘解确定观测数据的拟合函数

$$y(t,x^*) = y(t,x_1^*,\cdots,x_n^*),$$

我们称为观测数据的最小二乘拟合. 由于 $\gamma_i(x)$ 是 x 的线性函数，又称为线性最小二乘拟合.

8.2 超定方程组及其最小二乘解

8.2.1 超定方程组的最小二乘解

设

$$Ax = y, \qquad (8-12)$$

其中

$$A = (a_{ij})_{m\times n}, m > n,$$
$$x = (x_1,\cdots,x_n)^T,$$
$$y = (y_1,\cdots,y_m)^T.$$

因为方程的个数 m 大于未知量的个数 n，故称式(8－12)为超定方程组.

当 $\mathrm{rank}(A) = \mathrm{rank}(A|y)$ 时，式(8－12)相容，即存在 $x \in \mathbf{R}^n$ 使式(8－12)成立. 称 x 叫(8－12)的常义解.

当 $\mathrm{rank}(A) \neq \mathrm{rank}(A|y)$ 时，式(8－12)不相容，即对于 $\forall x \in \mathbf{R}^n$ 式(8－12)恒不成立. 在此情形式(8－12)不存在常义解.

例如超定方程组

$$\begin{cases} x_1 + x_2 = 3, \\ x_1 - x_2 = 1, \\ x_1 + 2x_2 = 4, \end{cases} \qquad (8-13)$$

其系数矩阵

$$A=\begin{bmatrix}1 & 1\\ 1 & -1\\ 1 & 2\end{bmatrix},x=(x_1,x_2)^{\mathrm{T}},b=(3,1,4)^{\mathrm{T}},\mathrm{rank}(A)=2,$$

而增广矩阵 $[A\mid b]=\begin{bmatrix}1 & 1 & 3\\ 1 & -1 & 1\\ 1 & 2 & 4\end{bmatrix}$， $\mathrm{rank}(A\mid b)=2$，

由于
$$\mathrm{rank}(A)=\mathrm{rank}(A\mid b),$$

故式(8－13)不仅有解，而且有唯一解

$$x^*=(2,1)^{\mathrm{T}}.$$

又如方程组

$$\begin{cases}x_1+x_2=3,\\ x_1-x_2=1,\\ x_1+2x_2=1,\end{cases}\tag{8－14}$$

$\mathrm{rank}(A)=2$，而 $\mathrm{rank}(A\mid b)=3$，故式(8－14)无常义解.

当式(8－12)无常义解时，令

$$\gamma(x)=Ax-y\neq 0,$$

解无约束极小问题

$$\min_{x\in\mathbf{R}^2}\|\gamma(x)\|_2^2,\tag{8－15}$$

其中
$$\gamma(x)=Ax-y=\sum_{i=1}^{m}\left[\sum_{j=1}^{n}a_{ij}x_j-y_i\right]^2.$$

极小问题式(8－15)的解 $x^*=(x_1^*,\cdots,x_n^*)^{\mathrm{T}}$ 称为超定方程组(8－12)的最小二乘解.且当 x^* 为式(8－12)的常义解时，则因 $\gamma(x^*)=0$，从而有 $\|\gamma(x^*)\|_2^2=0$.

下面我们进一步讨论由数据拟合形成的最小二乘问题式(8－11)与由超定方程求解所形成的最小二乘问题式(8－15)之间的联系.

在式(8－11)中，$\phi_j(t_i)$ 是 $\phi_j(t)$ 在 t_i 的值，而函数系 $\phi_1(t),\cdots,\phi_n(t)$ 分别在 $t=t_1,\cdots,t_m$ 处所取得的值形成矩阵 $[\phi_j(t_i)]_{\substack{i=1,\cdots,m\\ j=1,\cdots,n}}$

即
$$[\phi_j(t_i)]=\begin{bmatrix}\phi_1(t_1) & \cdots & \phi_j(t_1) & \cdots & \phi_n(t_1)\\ \vdots & & \vdots & & \vdots\\ \phi_1(t_i) & \cdots & \phi_j(t_i) & \cdots & \phi_n(t_i)\\ \vdots & & \vdots & & \vdots\\ \phi_1(t_m) & \cdots & \phi_j(t_m) & \cdots & \phi_n(t_m)\end{bmatrix}$$

若令 $\phi_j(t_i)=a_{ij}$，即有

$$[\phi_j(t_i)]_{m\times n}=[a_{ij}]_{m\times n}\xlongequal{记}\boldsymbol{A},$$

因此问题式(8－11)与式(8－15)的提法是等价的，因为

$$\|\gamma(\boldsymbol{x})\|_2^2=\sum_{i=1}^{m}\left[\sum_{j=1}^{n}x_j\phi_j(t_i)-y_i\right]^2$$
$$=\sum_{i=1}^{m}\left[\sum_{j=1}^{n}a_{ij}x_j-y_i\right]^2=\|\boldsymbol{Ax}-\boldsymbol{y}\|_2^2.$$

注意到，在数据拟合问题式(8－11)中

$$\gamma_i(\boldsymbol{x}^*)=\sum_{j=1}^{n}x_j^*\phi_j(t_i)-y_i\neq 0,\ i=1,2,\cdots,m,$$

即 $\boldsymbol{Ax}^*\neq\boldsymbol{y}$，$\gamma(\boldsymbol{x}^*)=\boldsymbol{Ax}^*-\boldsymbol{y}\neq 0$，因此，拟合问题所形成的超定方程无常义解.

例 8.1(多项式拟合)　设观测数据 (t_i,y_i) $i=1,2,\cdots,m$，取

$$\{\phi_j(t)\}=\{1,t,t^2,\cdots,t^{n-1}\},n<m,$$

构造关于 $a_i^*(i=1,\cdots,n)$ 的线性模型

$$y(t,\boldsymbol{x})=a_1^*+a_2^*t+\cdots+a_n^*t^{n-1},\qquad(8-16)$$

其中 $\boldsymbol{x}^*=(a_1^*,\cdots,a_n^*)^{\mathrm{T}}$，试确定式(8－16)中的 $a_i^*(i=1,\cdots,n)$.

若令

$$\boldsymbol{A}=\begin{bmatrix}1 & t_1 & t_1^2 & \cdots & t_1^{n-1}\\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & t_m & t_m^2 & \cdots & t_m^{n-1}\end{bmatrix},$$

$$\boldsymbol{y}=(y_1,\cdots,y_m)^{\mathrm{T}},$$

则问题变为求解如下线性最小二乘问题

$$\min_{x\in\mathbf{R}^n}\|\boldsymbol{Ax}-\boldsymbol{y}\|_2^2,\qquad(8-17)$$

式(8－17)的解为 $x^* = (a_1^*, \cdots, a_n^*)^T$.

例 8.2(三角函数拟合)　设(t_i, y_i) $i=1,2,\cdots,5$ 如表 8－2 所示：

表 8－2

t_i	0	1/4	$\pi/2$	$3\pi/4$	π
y_i	1	1/2	0	1/2	－1

构造线性模型.

$y(t, a^*, b^*) = a^* \sin t + b^* \cos t$,试确定待定参数 a^*, b^*

若令
$$A = \begin{bmatrix} \sin 0 & \cos 0 \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \\ \sin\frac{3\pi}{4} & \cos\frac{3\pi}{4} \\ \sin\pi & \cos\pi \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & 0 \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 0 & -1 \end{bmatrix},$$

$$y = \left(1, \frac{1}{2}, 0, \frac{1}{2}, -1\right)^T,$$

则问题变为求解如下线性最小二乘问题

$$\min_{x \in \mathbf{R}^2} \| Ax - y \|_2^2 \qquad (8－18)$$

式(8－18)的解为 $x^* = (a^*, b^*)^T$.

8.2.2　最小二乘解的存在性及唯一性

定理 8.1　超定方程组 $Ax = y$ 总存在最小二乘解,即总存在 $x^* \in \mathbf{R}^n$ 使

$$\min_{x \in \mathbf{R}^n} \| Ax - y \|_2^2 = \| Ax^* - y \|_2^2.$$

证　令 $R(A) = \{Z = Ax = \sum_{j=1}^{n} x_j a_j \mid A \in \mathbf{R}^{m\times n}, x \in \mathbf{R}^n\}$,　(8－19)

即 $R(A)$是 A 的所有列 $a_j (j=1,\cdots,n)$组成的线性子空间.

如图 8－2,据式(8－19),$R(A) \subset \mathbf{R}^m$,对 $\forall x \in \mathbf{R}^n, Ax \in R(A)$,令 $y = y_1 + y_2$,其中 y_1 是 y 在 $R(A)$ 上的投影,y_2 是 y 在 $R(A)$的正交补上的投影,欲证存在 x^*,使

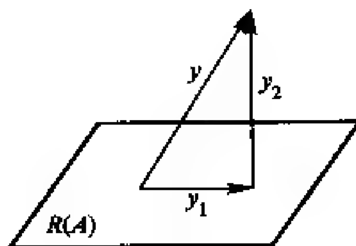


图 8－2

$$\|Ax^*-y\|_2^2<\|Ax-y\|_2^2 \quad 对\ \forall x\neq x^*\ 成立.$$

事实上对 $\forall x\in\mathbf{R}^n$，有

$$\|Ax-y\|_2^2=\|Ax-(y_1+y_2)\|_2^2=\|(Ax-y_1)-y_2\|_2^2$$

$$=\|Ax-y\|_2^2-2(Ax-y_1)^Ty_2+\|y_2\|_2^2.$$

因为 $y_1\in R(A)$，则 $Ax-y_1\in R(A)$，
所以 $(Ax-y_1)\perp y_2$，于是有

$$\|Ax-y\|_2^2=\|Ax-y_1\|_2^2+\|y_2\|_2^2.$$

又因为 $Ax\in R(A)$ 是 $R(A)$ 的泛元素，故由 $y_1\in R(A)$，总可以取到 x^*，使

$$Ax^*=y_1,$$

即 $\quad \|Ax^*-y\|_2^2=\|y_2\|_2^2<\|Ax-y\|_2^2 \quad 对\ \forall x\neq x^*\ 成立.$

定理 8.2　$x^*\in\mathbf{R}^n$ 成为如下线性最小二乘问题的解

$$\min_{x\in\mathbf{R}^n}\|Ax-y\|_2^2 \tag{8-20}$$

当且仅当 x^* 满足

$$A^T(Ax^*-y)=0,$$

即

$$A^TAx^*=A^Ty, \tag{8-21}$$

并且称式(8-21)是线性最小二乘问题式(8-20)的法方程.

证　充分性　设 x^* 满足式(8-21)，欲证

$$\min_{x\in\mathbf{R}^n}\|Ax-y\|_2^2=\|Ax^*-y\|_2^2,$$

只需证明对 $\forall x\in\mathbf{R}^n$ 使

$$\|Ax-y\|_2^2\geqslant\|Ax^*-y\|_2^2$$

事实上，令 $x=x^*+z$，则有

$$\|Ax-y\|_2^2=\|A(x^*+z)-y\|_2^2=\|(Ax^*-y)+Az\|_2^2$$

$$=\|Ax^*-y\|_2^2+2(Ax^*-y,Az)+\|Az\|_2^2$$

$$=\|Ax^*-y\|_2^2+2z^TA^T(Ax^*-y)+\|Az\|_2^2$$

$$= \| Ax^* - y \|_2^2 + \| Az \|_2^2$$

$$\geqslant \| Ax^* - y \|_2^2.$$

由 $x \in \mathbf{R}^n$ 的任意性知

$$\min_{x \in \mathbf{R}^n} \| Ax - y \|_2^2 = \| Ax^* - y \|_2^2.$$

必要性(反证法)　设 x^* 是极小问题 $\min\limits_{x \in \mathbf{R}^n} \| Ax - y \|_2^2$ 的解,但假定 $A^T(Ax^* - y) = z \neq 0$,令 $u = x^* - \varepsilon v (\varepsilon > 0)$,

$$\| Au - y \|_2^2 = \| (Ax^* - y) + \varepsilon Av \|_2^2$$

$$= \| Ax^* - y \|_2^2 - 2\varepsilon v^T A^T (Ax^* - y) + \varepsilon^2 \| Av \|_2^2$$

$$= \| Ax^* - y \|_2^2 - 2\varepsilon v^T z + \varepsilon^2 \| Av \|_2^2.$$

因为 $z \neq 0$,故对充分小的 $\varepsilon > 0$ 有

$$2\varepsilon v^T z > \varepsilon^2 \| Av \|_2^2,$$

所以

$$\| Au - y \|_2^2 \leqslant \| Ax^* - y \|_2^2,$$

这与假设矛盾.故有

$$A^T(Ax^* - y) = 0,$$

即式(8-21)成立.

推论 8.1(唯一性条件)　设线性最小二乘问题　$\min\limits_{x \in \mathbf{R}^n} \| Ax - y \|_2^2$

(8-22)

(1) 若 A^TA 非奇异,则式(8-22)有唯一解 $x = (A^TA)^{-1}A^Ty$;

(2) 若 A^TA 奇异,则式(8-22)有无穷多个解,可表为 $x + \alpha z$.其中 x 是(8-22)的任一解,z 满足 $Az = 0$,α 为任意实数.

证　若 A^TA 非奇异,则 $A^TAx = A^Ty$ 有唯一解

$$x = (A^TA)^{-1}A^Ty.$$

若 A^TA 奇异,则 A 列退化*,即 A 的列向量线性相关,故存在 $z \neq 0$ 使得

$$Az = 0.$$

注:设 $A \in \mathbf{R}^{m \times n}$,则有如下结论

(1) A^TA 非奇异的充要条件是 A 列满秩;

(2) A^TA 奇异的充要条件是 A 列退化.

设若 x 为式(8－20)的一个解，则 $x+\alpha z$ 必满足法方程

$$A^TA(x+\alpha z)=A^Ty.$$

事实上　$A^TA(x+\alpha z)=A^TAx+\alpha A^TAz=A^TAx+0=A^Ty$，

即 $x+\alpha z$ 亦为式(8－20)的解，由于 α 为任意实数，故(8－20)的解不唯一。

8.2.3　举例

例 8.3　设观测数据为 $(t_i,y_i)\, i=1,2,\cdots,m$，试用最小二乘法确定线性模型 $y(t,x)=a_1^*+a_2^*t+\cdots+a_n^*t^{n-1}$ 中的参数 $a_i^*\,(i=1,\cdots,n)$.

解　令

$$A=\begin{bmatrix}1 & t_1 & t_1^2 & \cdots & t_1^{n-1}\\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & t_m & t_m^2 & \cdots & t_m^{n-1}\end{bmatrix},$$

$$x^*=(a_1^*,\cdots,a_n^*)^T,$$

$$y=(y_1,\cdots,y_m)^T.$$

问题变成求解线性最小二乘问题

$$\min_{x\in\mathbf{R}^n}\|Ax-y\|_2^2,$$

由定理 8.2 直接解法方程(8－21)，即解方程组

$$A^TAx=A^Ty,$$

其中

$$A^TA=\begin{bmatrix}m & \sum t_i & \sum t_i^2 & \cdots & \sum t_i^{n-1}\\ \sum t_i & \sum t_i^2 & \sum t_i^3 & \cdots & \sum t_i^n\\ \sum t_i^2 & \sum t_i^3 & \sum t_i^4 & \cdots & \sum t_i^{n+1}\\ \vdots & \vdots & \vdots & & \vdots\\ \sum t_i^{n-1} & \sum t_i^n & \sum t_i^{n+1} & \cdots & \sum t_i^{2(n-1)}\end{bmatrix},$$

$$A^Ty=\left(\sum y_i,\sum t_iy_i,\cdots,\sum t_i^{n-1}y_i\right)^T.$$

矩阵 A^TA 及右端向量 A^Ty 中的所有元素可按表 8－3 进行计算：

表 8－3

i	t_i	y_i	t_iy_i	t_i^2	$t_i^2y_i$	…	t_i^{n-1}	$t_i^{n-1}y_i$	t_i^n	…	t_i^{2n-2}
1	t_1	y_1	t_1y_1	t_1^2	$t_1^2y_1$	…	t_1^{n-1}	$t_1^{n-1}y_1$	t_1^n	…	t_1^{2n-2}
2	t_2	y_2	t_2y_2	t_2^2	$t_2^2y_2$	…	t_2^{n-1}	$t_2^{n-1}y_2$	t_2^n	…	t_2^{2n-2}
⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮		⋮
m	t_m	y_m	t_my_m	t_m^2	$t_m^2y_m$	…	t_m^{n-1}	$t_m^{n-1}y_m$	t_m^n	…	t_m^{2n-2}
	$\sum t_i$	$\sum y_i$	$\sum t_iy_i$	$\sum t_i^2$	$\sum t_i^2y_i$	…	$\sum t_i^{n-1}$	$\sum t_i^{n-1}y_i$	$\sum t_i^n$	…	$\sum t_i^{2n-2}$

例 8.4　解超定方程组

$$\begin{cases} x_1 + x_2 = 3, \\ x_1 - x_2 = 1, \\ x_1 + 2x_2 = 4, \end{cases} \tag{8－23}$$

解　设 $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $b = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}$.

因为
$$\text{rank}(A) = \text{rank}(A \mid b) = 2,$$
所以方程组(8－23)相容，且 $\text{rank}(A) = 2$(未知量的个数)，故有唯一解，直接解式(8－23)得常义解

$$x_1 = 2,\quad x_2 = 1,$$

即
$$x = (x_1, x_2)^T = (2,1)^T.$$

现用最小二乘法求解式(8－23)，因为式(8－23)相容，故存在 x^* 使得

$$\| Ax^* - b \|_2^2 = \min_{x \in \mathbf{R}^2} \| Ax - b \|_2^2,$$

作法方程
$$A^TAx^* = A^Tb,$$

即
$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 10 \end{bmatrix},$$

又因为 A^TA 非奇异，故由推论知有唯一解.

$$x^* = (2,1)^T,$$

可见，在超定方程相容的情形，常义解与最小二乘解是一致的.

例 8.5　解超定方程组

$$\begin{cases} x_1 + x_2 = 3, \\ x_1 - x_2 = 1, \\ x_1 + 2x_2 = 1. \end{cases} \tag{8－24}$$

解　令 $A=\begin{bmatrix}1&1\\1&-1\\1&2\end{bmatrix},x=\begin{bmatrix}x_1\\x_2\end{bmatrix},b=\begin{bmatrix}3\\1\\1\end{bmatrix}$,

因为
$$\text{rank}(A)=2,\quad \text{rank}(A|b)=3,$$
所以方程组(8－24)不相容,只有最小二乘解.其法方程为
$$\begin{cases}3x_1+2x_2=5,\\2x_1+6x_2=4,\end{cases}$$
由此得式(8－24)的唯一解,$x=\left(\frac{11}{7},\frac{1}{7}\right)^{\mathrm T}$.

例 8.6　解超定方程组
$$\begin{cases}x_1+x_2=3,\\2x_1+2x_2=6,\\-x_1-x_2=1.\end{cases}\tag{8－25}$$

解　显然式(8－25)为矛盾方程组,只能有最小二乘解.

因为　$A^{\mathrm T}A=\begin{bmatrix}6&6\\6&6\end{bmatrix}$为奇异矩阵,

而
$$A^{\mathrm T}b=(14,14)^{\mathrm T}\neq 0,$$
解法方程　$A^{\mathrm T}A\begin{bmatrix}x_1\\x_2\end{bmatrix}=A^{\mathrm T}b$,得 $x_1+x_2=\frac{7}{3}$,

由此得式(8－25)的无穷多组最小二乘解
$$x_1=\frac{7}{3}-x_2\quad(x_2\text{ 为任意数}),$$
且对任意的 x_2,$\|Ax-b\|_2^2=\frac{40}{3}$.

例 8.7　解不定方程组
$$\begin{cases}x_1+x_2+x_3=6,\\2x_1+2x_2+2x_3=2.\end{cases}\tag{8－26}$$

解　式(8－26)为矛盾方程,无常义解,但可求最小二乘解.

由
$$A^{\mathrm T}A=\begin{bmatrix}5&5&5\\5&5&5\\5&5&5\end{bmatrix},\quad A^{\mathrm T}b=\begin{bmatrix}10\\10\\10\end{bmatrix},x=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}.$$

法方程
$$x_1+x_2+x_3=2,$$
任取 x_2,x_3 为自由未知量,得无穷多个解

$$x_1 = 2 - x_2 - x_3,$$

且对任意 x_2, x_3，$\| Ax - b \|_2^2 = 0$.

8.3　Golub 方法(用正交分解求最小二乘解)

为克服法方程 $A^TAx = A^Ty$ 出现病态所造成的困难，1965 年 Golub 和 Businger 提出避开法方程，直接用正交分解法解超定方程组(8－12).

8.3.1　矩阵的正交分解

引理 8.1(方阵的正交分解)　设 $A = (a_{ij})$ 为非奇异的 n 阶方阵，则存在正交阵 $Q_{n\times n}$ 及上三角阵 $R_{n\times n}$，使

$$QA = R, \tag{8-27}$$

其中 $Q = H_{n-1} \cdot H_{n-2} \cdot \cdots \cdot H_1$($H_i$ 为 n 阶 Householder 矩阵，附录 I 中将详细介绍该矩阵的定义和性质)，由于 $Q^T = Q^{-1}$，(8－27)也可写成

$$A = Q^TR. \tag{8-28}$$

(8－27)或(8－28)称为 A 的 QR 分解式.

引理 8.2(一般矩阵的 QU 分解)　设 $A = (a_{ij})_{m\times n}$，$m > n$，且 $\mathrm{rank}(A) = r < n$，则存在正交阵 $Q_{m\times m}$ 及上梯形阵 $U_{m\times n}$，使

$$Q_{m\times m}A_{m\times n} = U_{m\times n}, \tag{8-29}$$

其中 $Q_{m\times m} = H_\gamma \cdot H_{\gamma-1} \cdot \cdots \cdot H_1$($H_i$ 为 m 阶 Householder 矩阵)，而

$$U_{m\times n} = \left[\frac{U^*_{r\times n}}{O_{(m-r)\times n}}\right],$$

$$U^* = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1r} & u_{1,r+1} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2r} & u_{2,r+1} & \cdots & u_{2n} \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & & u_{rr} & u_{r,r+1} & \cdots & u_{rn} \end{bmatrix}.$$

设有超定方程

$$Ax = y, \tag{8-30}$$

若 $A = (a_{ij})_{m\times n}$ 为列满秩，即 $\mathrm{rank}(A) = n$，据引理 8.2，存在正交阵 $Q_{m\times m}$ 使

$$Q_{m\times m}A_{m\times n} = U_{m\times n} = \left[\frac{R_{n\times n}}{O_{(m-n)\times n}}\right],$$

其中

$$R_{n\times n} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}.$$

若令
$$Q_{m\times n}\ y = b = \left[\frac{p_{n\times 1}}{q_{(m-n)\times 1}}\right],$$
因此，用 Q 同时变换(8－30)的左右端得
$$QAx = Qy,$$
即
$$Ux = b,$$
即
$$\left[\frac{R_{n\times n}}{O}\right]x = \left[\frac{p}{q}\right]. \tag{8－31}$$

若(8－30)存在常义解，则经 Q 变换后(8－31)亦有常义解，于是有
$$Rx = p,\quad q = 0.$$

若(8－30)只有最小二乘解，我们将证明(8－31)亦有相同的最小二乘解，且为三角方程组
$$Rx = p \tag{8－32}$$
的解. 此时由于(8－31)并非真正等式，若(8－32)成立，则 $q\neq 0$.

定理 8.3　超定方程组(8－30)与三角方程组(8－32)同解.

证　首先证明(8－30)与(8－31)同最小二乘解. 事实上，由(8－31)有
$$Q(Ax-y) = Ux-b = \left[\frac{R}{O}\right]x-\left[\frac{p}{q}\right] = \left[\frac{Rx-p}{-q}\right],$$
而
$$\|Q(Ax-y)\|_2^2 = [Q(Ax-y)]^{\mathrm T}[Q(Ax-y)] = (Ax-y)^{\mathrm T}Q^{\mathrm T}Q(Ax-y)$$
$$=(Ax-y)^{\mathrm T}(Ax-y) = \|Ax-y\|_2^2,$$
显然，若 x^* 为$\|Ax-y\|_2^2$ 的极小点，必为$\|Q(Ax-y)\|_2^2$ 的极小点，反之亦真.

即是说(8－30)与(8－31)同最小二乘解. 下面进一步证明最小二乘解即为(8－32)的解.

因为
$$\|Q(Ax-y)\|_2^2 = (Rx-p,q)\left[\frac{Rx-p}{q}\right]$$
$$=(Rx-p)^{\mathrm T}(Rx-p)+q^{\mathrm T}q$$
$$=p^{\mathrm T}p+q^{\mathrm T}q+(Rx)^{\mathrm T}(Rx)-2p^{\mathrm T}(Rx), \tag{8－33}$$

所以　当 $p=Rx$ 时，(8－33)成为
$$\|Q(Ax-y)\|_2^2 = q^{\mathrm T}q, \tag{8－34}$$

而
$$p^Tp+(Rx)^T(Rx)-2p^T(Rx)$$
$$=\|p\|_2^2+\|Rx\|_2^2-2\|p\|_2\|Rx\|_2\cos(p,Rx)$$
$$\geqslant\|p\|_2^2+\|Rx\|_2^2-2\|p\|_2\|Rx\|_2$$
$$=(\|p\|_2-\|Rx\|_2)^2\geqslant 0,$$

所以 (8－34)达到极小，因此(8－32)的解即为$\|Q(Ax-y)\|_2^2$的极小点，亦为$\|Ax-y\|_2^2$的极小点，反之亦真.

至此，我们证明了(8－30)与(8－31)同最小二乘解，且为(8－32)的解.

定理8.3指示了借助矩阵正交分解，解超定方程组(8－30)的途径，其具体步骤归纳如下：

1. 用Householder矩阵同时变换A及y.

$$QA=(H_n,H_{n-1},\cdots,H_1)A=\left[\frac{R_{n\times n}}{O_{(m-n)\times n}}\right],$$

$$Qy=(H_n,H_{n-1},\cdots,H_1)y=\left[\frac{p_{n\times 1}}{q_{(m-n)\times 1}}\right]$$

(注意，若$Ax=y$存在常义解，则$q=0$，若$Ax=y$不存在常义解，则$q\neq 0$).

2. 解三角方程组　$Rx=p$，

其中
$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix},$$

$$p=(p_1,\cdots,p_n)^T,$$

求解公式为：

$$\begin{cases} x_n=\dfrac{p_n}{r_{nn}}, \\ x_i=\dfrac{p_i-\sum\limits_{j=i+1}^{n}x_j\gamma_{ij}}{r_{ii}}\ (i=n-1,\cdots,1), \end{cases} \tag{8－35}$$

$x=(x_1,\cdots,x_n)^T$即为(8－36)的最小二乘解.

8.3.2 Golub 算法

设 $Ax=y$, $A_{m\times n}$ 为列满秩,

记
$$A_0 = A = \begin{bmatrix} a_{11}^{(0)} & \cdots & a_{1n}^{(0)} \\ \vdots & & \vdots \\ a_{m1}^{(0)} & \cdots & a_{mn}^{(0)} \end{bmatrix},$$
$$b_0 = y = (b_{10}, \cdots, b_{m0})^{\mathrm{T}},$$
对 A_0 及 b_0 实行 n 次 Householder 变换,即
$$H_n \cdots H_1(A_0) = A_n = U,$$
$$H_n \cdots H_1(b_0) = b_n = b,$$
可归纳为如下递推形式
$$\begin{cases} A_k = H_k A_{k-1}, \\ b_k = H_k b_{k-1}, \end{cases} \quad k = 1,2,\cdots,n, \qquad (8-36)$$
第一次变换的结果是
$$\begin{cases} A_1 = H_1 A_0, \\ b_1 = H_1 b_0. \end{cases} \qquad (8-37)$$
式中 H_1 可利用 A_0 的第一列元素来构造
$$H_1 = 1 - \beta_1 v_1 v_1^{\mathrm{T}},$$
其中
$$\sigma_1 = \mathrm{sign}(a_{11}^{(0)}) \sqrt{\sum_{i=1}^{m} (a_{i1}^{(0)})^2},$$
$$\beta_1 = \frac{1}{\sigma_1(a_{11}^{(0)} + \sigma_1)},$$
$$v_1 = (a_{11}^{(0)} + \sigma_1, a_{21}^{(0)}, \cdots, a_{m1}^{(0)})^{\mathrm{T}},$$
于是(8-37)化为如下常用形式:
$$\begin{cases} A_1 = A_0 - v_1 q_1^{\mathrm{T}}, \\ b_1 = b_0 - v_1(p_1^{\mathrm{T}} b_0), \end{cases}$$
其中
$$q_1^{\mathrm{T}} = \beta_1 v_1^{\mathrm{T}} A_0. \quad p_1^{\mathrm{T}} = \beta_1 v_1^{\mathrm{T}}$$

类似地,第 k 次递推公式(8-36)的实用形式为:
$$\begin{cases} A_k = A_{k-1} - v_k q_k^{\mathrm{T}}, \\ b_k = b_{k-1} - v_k(p_k^{\mathrm{T}} b_{k-1}), \end{cases} \quad k = 1,2,\cdots,n, \qquad (8-38)$$
其中
$$\sigma_k = \mathrm{sign}(a_{kk}^{(k-1)}) \sqrt{\sum_{i=1}^{m} (a_{ik}^{(k-1)})^2},$$
$$\beta_k = \frac{1}{\sigma_k(a_{kk}^{(k-1)} + \sigma_k)},$$

$$v_k = (0,\cdots,a_{kk}^{(k-1)} + \sigma_k, a_{k+1,k}^{(k-1)},\cdots,a_{mk}^{(k-1)})^T,$$

$$q_k^T = \beta_k v_k^T A_{k-1}, p_1^T = \beta_k v_k^T$$

而 A_{k-1}, b_{k-1} 分别为

$$A_{k-1} = \left[\begin{array}{cccc|ccc} -\sigma_1 & a_{12}^{(0)} & \cdots & a_{1k-1}^{(0)} & a_{1k}^{(0)} & \cdots & a_{1n}^{(0)} \\ & -\sigma_2 & \cdots & a_{2k-1}^{(1)} & a_{2k}^{(1)} & \cdots & a_{2n}^{(1)} \\ & & \ddots & \vdots & \vdots & \vdots & \vdots \\ & & & -\sigma_{k-1} & a_{k-1,k}^{(k-1)} & \cdots & a_{k-1,n}^{(k-1)} \\ \hline & & & & a_{kk}^{(k-1)} & \cdots & a_{kn}^{(k-1)} \\ & O & & & \vdots & & \vdots \\ & & & & a_{mk}^{(k-1)} & \cdots & a_{mn}^{(k-1)} \end{array}\right],$$

$$b_{k-1} = (b_{1,k-1},\cdots,b_{m,k-1})^T,$$

最后结果是

$$A_n = H_n \cdots H_1 A_0 = U_{m\times n} = \left[\frac{R_{n\times n}}{O_{(m-n)\times n}}\right],$$

$$R_{n\times n} = \begin{bmatrix} -\sigma_1 & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ & -\sigma_2 & \cdots & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{nn}^{(n)} \end{bmatrix} \quad (记\ \sigma_i = a_{ii}^{(i)}, i = 1,\cdots,n-1),$$

$$b_n = H_n \cdots H_1 b_0 = (b_{1n},\cdots,b_{mn})^T,$$

综上所述,可得如下 Golub 基本算法

算法 8.1

① 给定 $A = (a_{ij})_{m\times n}, b = (b_1,\cdots,b_m)^T$ 及控制精度的误差限 ε;

② 置 $k = 1$;

③ 计算 $S_k = \sum_{i=1}^{m} a_{ik}^2$;

④ 计算 $\sigma_k = \mathrm{sign}(a_{kk})\sqrt{S_k}$;

⑤ 检查 $S_k < \varepsilon$?

是,停止计算(算法失败);

⑥ 计算 $\beta_k = \dfrac{1}{\sigma_k(a_{kk} + \sigma_k)}$

⑦ 计算 $v_k = (0,\cdots,0,a_{kk} + \sigma_k, a_{k+1,k},\cdots,a_{mk})^T$;

⑧ 计算 $p_k^T = \beta_k v_k^T, q_k^T = p_k^T A_{k-1}$;

⑨ 计算 $A_k = A_{k-1} - v_k q_k^T$,

$b_k = b_{k-1} - v_k(p_k^T b_{k-1})$;

⑩ 检查 $k < n$?

是,置 $k = k+1$,转③;

⑪ 令 $x_n = \frac{b_{nn}}{a_{nn}}$;

⑫ 对 $i = n-1,\cdots,1$ 计算

$$x_i = \frac{b_{in} - \sum_{j=i+1}^{n} a_{ij}^{(i)} x_j}{a_{ii}};$$

⑬ 打印 $x_i(i=1,\cdots,n)$,停.

为了减少舍入误差的影响,在算法 8.1 中还可增加"选主元"的步骤,考虑公式(8-38),

令
$$S_j^{(k)} = \sum_{i=k}^{n} (a^{(k-1)})^2, \quad j = k, k+1, \cdots, n,$$
$$S_l^{(k)} = \max_{k \le j \le n} \{S_j^{(k)}\},$$

取
$$\sigma_k = \mathrm{sign}(a_{kk}^{(k-1)}) \sqrt{S_l^{(k)}},$$

并且注意到这时要把 A_{k-1} 中的第 k 列和第 l 列进行交换,同时把 x 的相应分量也进行交换.

算法 8.2(选主元的 Golub 算法)

① 给定 $A = (a_{ij})_{m\times n}, b = (b_1,\cdots,b_m)^T$ 及控制精度的误差限 ε;

② 置 $k = 1$;

③ 对 $j = k,\cdots,n$ 计算 $S_j = \sum_{i=k}^{m} a_{ij}^{(2)}$;

④ 令 $S_l = \max_{k \le j \le n} \{S_j\}$;

⑤ 检查 $S_l < \varepsilon$?

是,停止计算(算法失败);

⑥ 检查 $l = k$?

是,转⑦;

否,交换 A 的第 k,l 两列,同时交换 x_k 和 x_l 的位置;

⑦ 计算 $\sigma_k = \mathrm{sign}(a_{kk})\sqrt{S_l}$;

⑧ 计算 $\beta_k = \frac{1}{\sigma_k(a_{kk} + \sigma_k)}$;

⑨ 计算 $v_k = (0,\cdots,0, a_{kk}+\sigma_k, a_{k+1,k},\cdots,a_{mk})^T$;

⑩ 计算 $p_k^T = \beta_k v_k^T, q_k^T = p_k^T A_{k-1}$;

⑪ 计算 $A_k = A_{k-1} - v_k q_k^T$,

$$b_k = b_{k-1} - v_k(q_k^T b_{k-1});$$

⑫ 检查 $k < n$?

是,置 $k = k+1$,转③;

⑬ 令 $x_n = \dfrac{b_{nn}}{a_{nn}}$;

⑭ 对 $i = n-1, \cdots, 1$,计算

$$x_i = \frac{b_{in} - \sum_{j=i+1}^{n} a_{ij}^{(i)} x_j}{a_{ii}};$$

⑮ 打印 $x_i (i = 1, \cdots, n)$,停.

例 8.8 解超定方程组

$$\begin{bmatrix} 1 & 1 \\ 0.01 & 0 \\ 0 & 0.01 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0.01 \\ 0.02 \end{bmatrix}.$$

显然,原方程组有常义解,$x = (1,2)^T$ 亦即最小二乘解. 这个解从理论上讲也可以通过解法方程组 $A^T A x = A^T y$ 求得

即解

$$\begin{bmatrix} 1.000\ 1 & 1 \\ 1 & 1.000\ 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3.000\ 1 \\ 3.000\ 2 \end{bmatrix}. \tag{8-39}$$

求得原超定方程组的解,但是由于 $\det(A^T A) = 0.000\ 1$ 很小,即(8-39)是一个病态方程组,求解该方程组极易受外界的干扰(如系数矩阵或右端向量的微小扰动)造成解的失真.

例如:若给右端向量一个微小扰动,令右端向量为 $(3.000\ 1, 3.000\ 1)^T$,则扰动后的解为 $x_1 = 1.5, x_2 = 1.5$;若给右端向量另一个微小扰动,令右端向量为 $(3, 3.000\ 2)^T$,则扰动后的解为 $x_1 = 0.5, x_2 = 2.5$.

下面我们用 Golub 算法求原超定方程组.

解 首先,求 $R_{2\times2}$ 及 p,即实现正交变换

$$H_2 H_1 A = \begin{bmatrix} R_{2\times2} \\ \hline O_{1\times2} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ 0 & r_{22} \\ \hline 0 & 0 \end{bmatrix},$$

$$H_2 H_1 y = \begin{bmatrix} p \\ \hline q \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \hline 0 \end{bmatrix}.$$

第一步

$$A_0 = A \xrightarrow{H_1} A_1 = H_1 A_0 = A_0 - v_1 q_1^T,$$

$$b_0 = b \xrightarrow{H_1} b_1 = H_1 b_0 = b_0 - v_1(p_1^T b_0).$$

① $\sigma_1 = \sqrt{1+0.01^2} = 1.000\ 0$;

② $\beta_1 = \dfrac{1}{\sigma_1(\sigma_1+1)} = 0.500\ 0$;

③ $v_1 = (2.000\ 0, 0.01, 0)^T$;

④ $p_1^T = (1.000\ 0, 0.005\ 0, 0)^T$;

⑤ $q_1^T = p_1^T A_0 = (1.000\ 0, 1.000\ 0)^T$;

⑥ $v_1 q_1^T = \begin{bmatrix} 2.000\ 0 & 2.000\ 0 \\ 0.01 & 0.01 \\ 0 & 0 \end{bmatrix}$,

$$A_1 = A_0 - v_1^T q_1 = \begin{bmatrix} -1.000\ 0 & -1.000\ 0 \\ 0 & -0.01 \\ 0 & 0.01 \end{bmatrix},$$

$b_1 = b_0 - v_1(p_1^T b_0) = (-3, -0.02, 0.02)^T$;

第二步

$$A_1 \xrightarrow{H_2} A_2 = H_2 A_1 = A_1 - v_2 q_2^T,$$

$$b_1 \xrightarrow{H_2} b_2 = H_2 b_1 = b_1 - v_2(p_2^T b_1).$$

① $\sigma_2 = -\sqrt{0.01^2+0.01^2} = -0.014\ 14$;

② $\beta_2 = \dfrac{1}{\sigma_2(\sigma_2-0.01)} = 2\ 929.633\ 7$;

③ $v_2 = (0, -0.024\ 14, 0.01)^T$;

④ $p_2^T = (0, -70.721\ 4, 29.296\ 3)^T$;

⑤ $q_2^T = p_2^T A_1 = (0, 1.000\ 1)^T$;

⑥ $v_2 q_2^T = \begin{bmatrix} 0 & 0 \\ 0 & -0.024\ 14 \\ 0 & 0.01 \end{bmatrix}$;

$$A_2 = A_1 - v_2 q_2^T = \begin{bmatrix} -1 & -1 \\ 0 & 0.014\ 14 \\ 0 & 0 \end{bmatrix};$$

$b_2 = b_1 - v_1(q_2^T b_1) = (-3, 0.028\ 3, 0)^T$;

由此得

$$R_{2\times2} = \begin{bmatrix} -1 & -1 \\ 0 & 0.014\ 14 \end{bmatrix},$$

$$p=(-3,0.0283)^T.$$

其次，解 $Rx=p$，据(8-35)得

$$\begin{cases}x_1=0.9986,\\x_2=2.0014,\end{cases}$$

即 $x^*=(0.9986,2.0014)^T$ 为原方程的最小二乘解.

附录 I　初等反射矩阵(H 矩阵)及其性质

（一）定义 8.1　设 $u \in \mathbf{R}^n$ 且 $\|u\|_2 = 1$，I 为 n 阶单位矩阵，称 $H = I - 2uu^T$ 叫做 Householder 矩阵(简称 H 矩阵)．由 H 矩阵确定的变换叫 H 变换．

该定义的更一般的形式是：

设 $w \in \mathbf{R}^n$，$w \neq 0$，称 $I - 2\dfrac{ww^T}{\|w\|_2^2} = I - \beta ww^T$ 叫做 H 矩阵．其中 $\beta = \dfrac{2}{\|w\|_2^2}$．

（二）基本性质

1．对称性．即 $H^T = H$．

事实上，$H^T = (I - 2uu^T)^T = I - 2(uu^T)^T = I - 2uu^T = H$．

2．正交性．即 $H^T H = I$．

事实上，
$$H^T H = (I - 2uu^T)^2 = I - 4uu^T + 4uu^T \cdot uu^T$$
$$= I - 4uu^T + 4u(u^T - u)u^T = I - 4uu^T + 4uu^T = I.$$

3．对合性．即 $H^2 = I$．

事实上，$H^2 = H \cdot H = H^T H = I$．

推广结论：$H^n = \begin{cases} H, & n \text{ 为奇}, \\ I, & n \text{ 为偶}. \end{cases}$

（三）几何意义

设 $v \in \mathbf{R}^n$，令 $Hv = v'$，可以证明 v' 恰为 v 在镜(线) l 下的镜像(如图 I －1)．

设 S 为过原点以向量 u 为法向量的超平面，即 $u^T x = 0$．

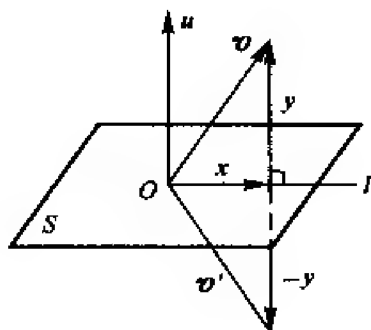
设 $v \in \mathbf{R}^n$，则有 $v = x + y$，其中 $x \in S$，$y \in S^{\perp}$，由

$$Hx = (I - 2uu^T)x = x - 2uu^T x = x,$$

$$Hy = (I - 2uu^T)y = y - 2uu^T y = y - 2uu^T(\lambda u)\ (y \text{ 与 } u \text{ 共线})$$
$$= y - 2\lambda u = y - 2y = -y,$$

所以
$$Hv = Hx + Hy = x - y = v'.$$

可见，对于给定的 v，经 H 作用后，$Hv = v'$ 恰为 v 在镜(线) l 下的镜像．



图Ⅰ-1

（四）定理（不予证明）

定理 8.4　设 $\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$，且 $\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2$，则总存在 $\boldsymbol{H}$ 阵，使得 $\boldsymbol{Hx}=\boldsymbol{y}$，其中

$$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}},\quad \boldsymbol{u}=\frac{\boldsymbol{y}-\boldsymbol{x}}{\|\boldsymbol{y}-\boldsymbol{x}\|_2}.$$

例 8.9　设 $\boldsymbol{x}=(1,2,1)^{\mathrm{T}},\boldsymbol{y}=(-1,2,-1)^{\mathrm{T}}$，求 $\boldsymbol{H}$ 使得 $\boldsymbol{Hx}=\boldsymbol{y}$.

解　$\boldsymbol{y}-\boldsymbol{x}=(-2,0,-2)^{\mathrm{T}},\|\boldsymbol{y}-\boldsymbol{x}\|_2=2\sqrt{2}$，

据定理 8.4，取 $\boldsymbol{u}=\dfrac{\boldsymbol{y}-\boldsymbol{x}}{\|\boldsymbol{y}-\boldsymbol{x}\|_2}=\left(-\dfrac{1}{\sqrt{2}},0,-\dfrac{1}{\sqrt{2}}\right)^{\mathrm{T}}$，

$$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}=\begin{bmatrix}0&0&-1\\0&1&0\\-1&0&0\end{bmatrix},$$

$$\boldsymbol{Hx}=\begin{bmatrix}0&0&-1\\0&1&0\\-1&0&0\end{bmatrix}\begin{bmatrix}1\\2\\1\end{bmatrix}=\begin{bmatrix}-1\\2\\-1\end{bmatrix}=\boldsymbol{y}.$$

定理 8.5　设 $\boldsymbol{x}\in\mathbf{R}^n,\boldsymbol{x}\neq 0$，则总存在 $\boldsymbol{H}$ 阵，使得

$$\boldsymbol{Hx}=-\sigma_1\boldsymbol{e}_1=(-\sigma_1,0,\cdots,0)^{\mathrm{T}},$$

其中

$$\begin{cases}\boldsymbol{H}=\boldsymbol{I}-\beta\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}},\\ \boldsymbol{v}=\boldsymbol{x}+\sigma_1\boldsymbol{e}_1,\\ \beta=\dfrac{1}{\sigma_1(\sigma_1+x_1)},\\ \sigma_1=\pm\|\boldsymbol{x}\|_2.\end{cases}\tag{8-40}$$

为减少舍入误差，通常取 $\sigma=\operatorname{sign}(x_1)\|\boldsymbol{x}\|_2$（$x_1$ 为 $\boldsymbol{x}$ 的第一分量），故式(8-40)可改写为

$$\begin{cases} H = I - \beta vv^T, \\ v_1 = x_1 + \sigma_1, \\ v_k = x_k (k = 2,3,\cdots,n), \\ \beta = \dfrac{1}{\sigma_1(\sigma_1 + x_1)} = \dfrac{1}{\sigma v_1}, \\ \sigma_1 = \text{sign}(x_1)\|x\|_2. \end{cases} \tag{8-41}$$

例 8.10　设 $x = (1,2,1)^T$，求 H 使得 $Hx = -\sigma_1 e_1$.

解　据定理 8.5 之式(8-41)

$$\sigma_1 = \text{sign}(1)\|x\|_2 = \sqrt{6},$$

$$v = (1+\sqrt{6},2,1)^T,$$

$$\beta = \frac{1}{\sigma_1 v_1} = \frac{1}{\sqrt{6}(1+\sqrt{6})} = \frac{1}{6+\sqrt{6}},$$

所以

$$H = I - \beta vv^T = \begin{bmatrix} -\dfrac{1}{\sqrt{6}} & -\dfrac{2}{\sqrt{6}} & -\dfrac{1}{\sqrt{6}} \\ -\dfrac{2}{\sqrt{6}} & \dfrac{2+\sqrt{6}}{6+\sqrt{6}} & -\dfrac{2}{6+\sqrt{6}} \\ -\dfrac{1}{\sqrt{6}} & -\dfrac{2}{6+\sqrt{6}} & \dfrac{5+\sqrt{6}}{6+\sqrt{6}} \end{bmatrix},$$

验证

$$Hx = -\sigma_1 e_1 = (-\sqrt{6},0,0)^T.$$

推论 8.2　设 $x \in \mathbf{R}^n$，$x \neq 0$，则总存在 H 矩阵，使得

$$Hx = (x_1,\cdots,x_{i-1},-\sigma_i,0,\cdots,0)^T,$$

其中

$$\begin{cases} H = I - \beta_i vv^T, \\ v_1 = v_2 = \cdots = v_{i-1} = 0, \\ v_i = x_i + \sigma_i, \\ v_k = x_k, k = i+1,\cdots,n, \\ \sigma_i = \text{sign}(x_i)\left(\sum\limits_{k=i}^{n} x_k^2\right)^{\frac{1}{2}}, \\ \beta_i = \dfrac{1}{\sigma_i(\sigma_i + x_i)} = \dfrac{1}{\sigma_i v_i}. \end{cases} \tag{8-42}$$

例 8.11　设 $x = (1,2,1)^T$，求 H 使得 $Hx = (1,-\sigma_2,0)^T$.

解　据推论 8.2 之式(8-42)

$$\sigma_2 = \text{sign}(2)\sqrt{2^2+1^2} = \sqrt{5},$$

$$v = (0,2+\sqrt{5},1)^T,$$

$$\beta_2 = \frac{1}{\sigma_2 v_2} = \frac{1}{\sqrt{5}(2+\sqrt{5})} = \frac{1}{5+2\sqrt{5}},$$

所以
$$H = I - \beta_2 vv^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-\dfrac{9+4\sqrt{5}}{5+2\sqrt{5}} & -\dfrac{5+\sqrt{5}}{5+2\sqrt{5}} \\ 0 & -\dfrac{5+\sqrt{5}}{5+\sqrt{5}} & 1-\dfrac{1}{5+2\sqrt{5}} \end{bmatrix},$$

验证
$$Hx = (1 \quad -\sqrt{5}, \quad 0)^{\mathrm{T}}.$$

（五）Householder 变换（H 变换）

设 $A = (a_{ij}^{(0)})_{m \times n}$，且 $\mathrm{rank}(A) = n$，试确定 n 个 H 矩阵 $H_i (i = 1, \cdots, n)$，使得

$$H_n H_{n-1} \cdots H_1 A_0 = \left[\frac{R_{n \times n}}{O_{(m-n)\times n}}\right],$$

其中
$$R_{n \times n} = \begin{bmatrix} -\sigma_1 & a_{12}^{(0)} & \cdots & a_{1n}^{(0)} \\ & -\sigma_2 & \cdots & a_{2n}^{(1)} \\ & & \ddots & \vdots \\ 0 & & & -\sigma_n \end{bmatrix}.$$

计算步骤如下：

第 1 步 $A_0 \xrightarrow{H_1} H_1 A_0 = A_1$，

$H_1 = I - \beta_1 v_1 v_1^{\mathrm{T}}$，其中

$$\begin{cases} \sigma_1 = \mathrm{sign}\,(a_{11}^{(0)}) \sqrt{\displaystyle\sum_{i=1}^{m} (a_{i1}^{(0)})^2}, \\ \beta_1 = \dfrac{1}{\sigma_1 (a_{11}^{(0)} + \sigma_1)}, \\ v_1 = (a_{11}^{(0)} + \sigma_1, a_{21}^{(0)}, \cdots, a_{m1}^{(0)})^{\mathrm{T}}. \end{cases}$$

$$A_1 = \begin{bmatrix} -\sigma_1 & a_{12}^{(0)} & \cdots & a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{m2}^{(1)} & \cdots & a_{mn}^{(1)} \end{bmatrix}.$$

第 k 步 $A_{k-1} \xrightarrow{H_k} H_k A_{k-1} = A_k$，

$H_1 = I - \beta_1 v_1 v_1^{\mathrm{T}}$，其中

$$
\begin{cases}
\sigma_k = \operatorname{sign}(a_{kk}^{(k-1)})\sqrt{\sum\limits_{i=1}^{m}(a_{ik}^{(k-1)})^2}, \\
\beta_k = \dfrac{1}{\sigma_k(a_{kk}^{(k-1)} + \sigma_k)}, \\
v_1 = (0,\cdots,0,a_{kk}^{(k-1)} + \sigma_k, a_{k+1,k}^{(k-1)},\cdots,a_{mk}^{(k-1)})^T.
\end{cases}
$$

$$
A_{k-1} = \begin{bmatrix}
-\sigma_1 & a_{12}^{(0)} & \cdots & a_{1k-1}^{(0)} & a_{1k}^{(0)} & \cdots & a_{1n}^{(0)} \\
 & -\sigma_2 & \cdots & a_{2k-1}^{(1)} & a_{2k}^{(1)} & \cdots & a_{2n}^{(1)} \\
 & & \ddots & \vdots & \vdots & & \vdots \\
 & & & -\sigma_{k-1} & a_{k-1,k}^{(k-1)} & \cdots & a_{k-1,n}^{(k-1)} \\
 & & & & a_{kk}^{(k-1)} & \cdots & a_{kn}^{(k-1)} \\
 & & & & \vdots & & \vdots \\
 & & & & a_{mk}^{(k-1)} & \cdots & a_{mn}^{(k-1)}
\end{bmatrix},
$$

$$
A_k = \begin{bmatrix}
-\sigma_1 & a_{12}^{(0)} & \cdots & a_{1k}^{(0)} & a_{1k+1}^{(0)} & \cdots & a_{1n}^{(0)} \\
 & -\sigma_2 & \cdots & a_{2k}^{(1)} & a_{2k+1}^{(1)} & \cdots & a_{2n}^{(1)} \\
 & & \ddots & \vdots & \vdots & & \vdots \\
 & & & -\sigma_k & a_{k,k+1}^{(k)} & \cdots & a_{kn}^{(k)} \\
 & & & & a_{k+1,k+1}^{(k+1)} & \cdots & a_{k+1,n}^{(k+1)} \\
 & & & & \vdots & & \vdots \\
 & & & & a_{m,k+1}^{(k+1)} & \cdots & a_{mn}^{(k+1)}
\end{bmatrix}.
$$

令 $k=1,2,\cdots,n$，得

$$
H_nH_{n-1}\cdots H_1A_0 = A_n,
$$

$$
A_n = \left[\frac{R_{n\times n}}{O_{(m-n)\times n}}\right],
$$

$$
R_{n\times n} = \begin{bmatrix}
-\sigma_1 & a_{12}^{(0)} & \cdots & a_{1n}^{(0)} \\
 & -\sigma_2 & \cdots & a_{2n}^{(1)} \\
 & & \ddots & \vdots \\
 & & & -\sigma_n
\end{bmatrix}.
$$

第9章　非线性最小二乘法

9.1　非线性最小二乘法问题

9.1.1　问题的提出

给出 $f_i(x)$ $(i=1,\cdots,m)$，设 $f_i(x)$ 为 $x\in\mathbf{R}^n$ 的非线性函数，令

$$f(x)=(f_1(x),\cdots,f_m(x))^{\mathrm{T}}.$$

$f(x)$ 为 x 的向量值函数，则

$$S(x)=\sum_{i=1}^{m}f_i^2(x)=f(x)^{\mathrm{T}}f(x)=\|f(x)\|_2^2$$

为 x 的平方和函数.

无约束极小问题

$$\min_{x\in\mathbf{R}^n}S(x) \tag{9-1}$$

叫做非线性最小二乘问题. 或者叫无约束极小平方和函数问题.

若 $f_i(x)$ $(i=1,\cdots,m)$ 为 x 的线性函数，即

$$f_i(x)=a_i^{\mathrm{T}}x+b_i，其中\ a_i\in\mathbf{R}^n,\ b_i\in\mathbf{R},$$

则有

$$f(x)=Ax+b,$$

其中

$$A=\begin{bmatrix}a_1^{\mathrm{T}}\\ \vdots\\ a_m^{\mathrm{T}}\end{bmatrix},\quad b=\begin{bmatrix}b_1\\ \vdots\\ b_m\end{bmatrix},$$

故

$$\|f(x)\|_2^2=\|Ax+b\|_2^2.$$

此时，问题(9-1)成为线性最小二乘问题.

9.1.2　问题的形式

1. 非线性模型的参数识别

设有 m 组观测数据

$$(t_i,\ y_i)\quad i=1,\cdots,m$$

其中

$$t_i\neq t_j\quad(i\neq j),$$

$$y_i = f(t_i) \quad (i = 1,\cdots,m).$$

令 $t_i = (t_{1i}, t_{2i}, \cdots, t_{li})$，将 m 组观测数据 $(t_i, y_i)_{i=1,2,\cdots,m}$ 列表如下：

表 9－1

t	t_1	$\cdots$	t_i	$\cdots$	t_m
y	y_1	$\cdots$	y_i	$\cdots$	y_m

此外还给出函数 $y = f(t)$ 的含参数 $x_i (i = 1,\cdots,n)$ 经验模型：

$$y = (t;\ x_1, \cdots, x_n), \tag{9-2}$$

其中 $x_1,\cdots,x_n$ 为待定参数，(9－2)式还可简记为

$$y(t, x),$$

其中 $x = (x_1,\cdots,x_n)^{\mathrm{T}}$，显然，当 t 取定观测值时，则

$$y(t_i, x)$$

为 x 的非线性函数.

现在的问题是确定 x 使得 $y(t,x)$ 在最小平方逼近意义下拟合于数据

$$(t_i,\ y_i),\ i = 1,\cdots,m,$$

即求解极小化问题

$$\min_{x\in\mathbb{R}^n} S(x) = \sum_{i=1}^{m} (y(t_i, x) - y_i)^2. \tag{9-3}$$

若令 $f_i(x) = y(t_i, x) - y_i (i = 1,\cdots,m)$ 为第 i 组数据的残差，则可令

$$f(x) = (f_1(x),\cdots,f_m(x))^{\mathrm{T}}$$

表示全部数据的残差向量，此时有

$$S(x) = \sum_{i=1}^{m} f_i^2(x) = \|f(x)\|_2^2. \tag{9-4}$$

同线性最小二乘问题比较：

(1) $y(t_i, x)$ 为 x 的非线性函数，因此它在一般情况下不能表为独立函数系 $\{\varphi_j(t)\}_{j=1,\cdots,n}$ 的线性组合；

(2) $y(t_i, x)$ 通常是从具体问题出发，通过经验选型而得到的；

(3) 对于 $S(x)$ 的极小点 x^*，要求 $f_i(x^*) \neq 0$，而对 $\forall x \in U(x^*;\varepsilon)$，要求 $f_i(x) \approx 0$.

令　　$$f_i(\boldsymbol{x}) = V(\boldsymbol{x},T_i) - V_i, i = 1,\cdots,22,$$

则有　　$$S(\boldsymbol{x}) = \sum_{i=1}^{22} f_i^2(\boldsymbol{x}),$$

问题成为求 $\boldsymbol{x}^*$ 使 $S(\boldsymbol{x}^*) = \min\limits_{\boldsymbol{x}\in\mathbf{R}^6} S(\boldsymbol{x})$.

2. 解非线性方程组

设有非线性方程组

$$\begin{cases} f_1(x_1,\cdots,x_n) = 0, \\ \cdots\cdots\cdots\cdots \\ f_m(x_1,\cdots,x_n) = 0. \end{cases} \tag{9-5}$$

其中 $m \geqslant n$,若 $m = n$,可用 Newton-Rephon 方法求解,这里讨论 $m > n$ 的情形.

令　　$$\boldsymbol{x} = (x_1,\cdots,x_n)^{\mathrm{T}},$$

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x}))^{\mathrm{T}},$$

则(9-5)可写成

$$\boldsymbol{f}(\boldsymbol{x}) = 0.$$

令　　$$S(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}) = \|\boldsymbol{f}(\boldsymbol{x})\|_2^2.$$

不难证明如下结论:求 $\boldsymbol{x}^*$ 使 $\boldsymbol{f}(\boldsymbol{x}^*) = 0$. 等价于求 $\boldsymbol{x}^*$ 使

$$S(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\mathbf{R}^n} S(\boldsymbol{x}),\text{且 } S(\boldsymbol{x}^*) = \mathbf{0}. \tag{9-6}$$

事实上,若存在 $\boldsymbol{x}^*$ 使 $\boldsymbol{f}(\boldsymbol{x}^*) = \mathbf{0}$,则有 $\|\boldsymbol{f}(\boldsymbol{x}^*)\|_2^2 = \mathbf{0}$,即 $S(\boldsymbol{x}^*) = \mathbf{0}$,但对 $\forall \boldsymbol{x} \in \mathbf{R}^n$,有 $S(\boldsymbol{x}) = \|\boldsymbol{f}(\boldsymbol{x})\|_2^2 \geqslant 0$,故得

$$S(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\mathbf{R}^n} S(\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{f}(\boldsymbol{x})\|_2^2,$$

即 $\boldsymbol{x}^*$ 是 $S(\boldsymbol{x})$ 的极小点.

反之,若存在 $\boldsymbol{x}^*$ 使 $S(\boldsymbol{x}^*) = \mathbf{0}$,则由 $\|\boldsymbol{f}(\boldsymbol{x}^*)\|_2^2 = \mathbf{0}$,必有 $\boldsymbol{f}(\boldsymbol{x}^*) = \mathbf{0}$.

注意两类极小问题的区别:对于参数识别问题(9-3),由于 $f_i(\boldsymbol{x}^*) \approx 0(i = 1,\cdots,m)$,故 $S(\boldsymbol{x}^*) \neq \mathbf{0}$,此类问题称为小残量问题;而对于非线性方程组求解所形成的问题(9-6),则因 $f_i(\boldsymbol{x}^*) = 0(i = 1,\cdots,m)$,故 $S(\boldsymbol{x}^*) = \mathbf{0}$,此类问题称为零残量问题.

9.1.3　解法概述

1. 古典分析方法

对于问题(9-3)或(9-6),令

$$\nabla S(\boldsymbol{x}) = \mathbf{0},$$

由于　　$$\nabla S(\boldsymbol{x}) = 2D\boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}),$$

$$D\boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}) = 0,$$

即 $$[\nabla f_1(\boldsymbol{x}),\cdots,\nabla f_m(\boldsymbol{x})]f(\boldsymbol{x})=0,$$

亦即 $$\sum_{i=1}^{m} f_i(\boldsymbol{x})\frac{\partial f_i(\boldsymbol{x})}{\partial x_j}=0,\quad j=1,\cdots,n.$$

令 $$\varphi_j(\boldsymbol{x})=\sum_{i=1}^{m} f_i(\boldsymbol{x})\frac{\partial f_i(\boldsymbol{x})}{\partial x_j},$$

有 $$\varphi_j(\boldsymbol{x})=0,\quad j=1,\cdots,n. \tag{9-7}$$

显然,通过解非线性方程组(9-7)来求解问题(9-3)或(9-6)都是不适宜的.

2. 下降算法

如前所述,下降算法是一种迭代方法,对于极小化问题(9-1),设 $\boldsymbol{x}^* \in \mathbf{R}^n$ 为极小点,从 $\boldsymbol{x}_0 \in \mathbf{R}^n$ 出发,按下述步骤进行迭代:

(1) 取定下降方向 $\boldsymbol{p}_0$: $\nabla S(\boldsymbol{x}_0)^{\mathrm{T}}\boldsymbol{p}_0<0$;

(2) 确定步长因子 α_0:即以 $\boldsymbol{p}_0$ 为方向作射线 $\boldsymbol{x}_0+\alpha\boldsymbol{p}_0$,在此射线上定出 $\boldsymbol{x}_1=\boldsymbol{x}_0+\alpha_0\boldsymbol{p}_0$ 使得 $S(\boldsymbol{x}_0+\alpha_0\boldsymbol{p}_0)<S(\boldsymbol{x}_0)$;

(3) 迭代:以 $\boldsymbol{x}_1=\boldsymbol{x}_0+\alpha_0\boldsymbol{p}_0$ 作为 $\boldsymbol{x}^*$ 的第一次近似值.

上述过程重复进行直至满足收敛准则. 可以归纳成如下算法:

初始点: $\boldsymbol{x}_0 \in \mathbf{R}^n$;

迭代公式: $$\begin{cases}\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k\\ \boldsymbol{p}_k:\ \nabla S(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{p}_k<0,\\ \alpha_k:\ S(\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k)<S(\boldsymbol{x}_k),\\ k=0,1,\cdots;\end{cases}$$

终止法则:若 $\boldsymbol{x}_k$ 满足指定精度就停止迭代,并且以 $\boldsymbol{x}_k$ 作为 $\boldsymbol{x}^*$ 的近似值,否则继续迭代,直到满足指定精度.

9.2 Gauss-Newton 算法(简称 G-N 算法)

9.2.1 G-N 方向的构造

对于极小问题 $$\min_{\boldsymbol{x}\in\mathbf{R}^n} S(\boldsymbol{x})=\frac{1}{2}\|f(\boldsymbol{x})\|_2^2, \tag{9-8}$$

G-N 方法按如下迭代公式计算:

给定初始近似 $\boldsymbol{x}^0$,令 $k=0,1,\cdots$,计算

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{p}_k,$$

其中 $$\boldsymbol{p}_k=-(\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k)^{-1}\nabla S(\boldsymbol{x}_k), \tag{9-9}$$

$$A_k = Df(x_k) = \left[\frac{\partial f_i(x)}{\partial x_j}\right]_{x=x_k} \quad (i = 1,2,\cdots,m;\ j = 1,2,\cdots,n),$$

$$\nabla S(x_k) = A_k^T f_k.$$

在(9－9)中，G-N 方向 p_k 可以看作是 Newton 方向的近似. 事实上，

$$p_k^{\text{Newton}} = -[\nabla^2 S(x_k)]^{-1} \nabla S(x_k),$$

而

$$\nabla^2 S(x_k) = Df(x_k)^T Df(x_k) + \sum_{i=1}^m f_i(x_k) \nabla^2 f_i(x_k)$$

$$= A_k^T A_k + \sum_{i=1}^m f_i(x_k) \nabla^2 f_i(x_k).$$

当其 $f_i(x) \approx 0$，对 $\forall x \in \cup (x^*; \varepsilon)$

$$p_k^{\text{Newton}} \approx -(A_i^T A_k)^{-1} \nabla S(x_k) \nabla S(x_k) = p_k^{\text{G-N}}.$$

另方面，G-N 方向也可以通过线性化的手段而得到. 首先，将 $f(x)$ 在 x_k 处线性化，即将 $f_i(x)$ 在 x_k 处线性展开：

$$f_i(x) = f_i(x_k + \Delta x_k) \approx f_i(x_k) + \nabla f_i(x_k)^T \Delta x_k, \quad i = 1,2,\cdots,m,$$

于是有

$$f(x) = f(x_k + \Delta x_k) \approx f(x_k) + \begin{bmatrix} \nabla f_1(x_k)^T \\ \vdots \\ \nabla f_m(x_k)^T \end{bmatrix} \Delta x_k = f(x_k) + A_k \Delta x_k,$$

其次解线性最小二乘问题：

$$\min \|A_k \Delta x_k + f_k\|_2^2, \tag{9－10}$$

其中

$$\Delta_k = x - x_k,$$

$$f_k = (f_1(x_k), \cdots, f_m(x_k))^T.$$

问题(9－10)的法方程为

$$A_k^T A_k \Delta x_k + A_k^T f_k = 0. \tag{9－11}$$

(9－11)的解即为(9－10)的极小解

$$\Delta x_k = -(A_k^T A_k)^{-1} A_k^T f_k,$$

此即为 G-N 方向.

9.2.2 G-N 算法及其局部收敛性质

算法 9.1

给出初始点 x_0 及指定精度 $\varepsilon_1, \varepsilon_2, \varepsilon_3$.

① 置 $k = 0$；

② 计算 $f_k = f(x_k)$，$S_k = S(x_k)$；

③ 计算 $A_k = Df(x_k) = \left[\frac{\partial f_i(x)}{\partial x_j}\right]_{x=x_k}$ $(i = 1,2,\cdots,m;\ j = 1,2,\cdots,n)$

及 $A_k^T A_k$;

④ 计算 $g_k = \nabla S(x_k) = A_k^T f_k$;

⑤ 解方程 $A_k^T A_k \Delta x_k = -g_k$;

⑥ 计算 $x_{k+1} = x_k + \Delta x_k$;

⑦ 计算 $f_{k+1} = f(x_{k+1}), S_{k+1} = S(x_{k+1})$;

⑧ 检查 H − 终止准则.

若满足,则输出 x_{k+1}, S_{k+1} 停止计算;

否则,$k \Leftarrow k+1, x_k \Leftarrow x_{k+1}, S_k \Leftarrow S_{k+1}$,转③.

下面将证明在一定条件下,由 G-N 法所产生的点列 $\{x_k\}$ 收敛于 $S(x) = f(x)^T f(x)$ 的稳定点 x^*,并且对于零向量问题至少有二阶收敛速率.

定理 9.1 设

(1) A, B 均为 n 阶方阵;

(2) A 非奇异且 $\|A^{-1}\| \leqslant \alpha$;

(3) $\|B - A\| \leqslant \beta$;

(4) $\alpha \cdot \beta < 1$,

则 B 非奇异且 $\|B^{-1}\| \leqslant \dfrac{\alpha}{1 - \alpha\beta}$. (9 − 12)

证 首先证明 $B^{-1} = \sum_{k=0}^{\infty} (I - A^{-1}B)^k A^{-1}$ (9 − 13)

事实上,因 $(I - A^{-1}B)^k = (A^{-1}(A - B))^k$,

而
$$\|A^{-1}(A - B)\| \leqslant \|A^{-1}\| \cdot \|A - B\| < \alpha \cdot \beta < 1,$$

故
$$\sum_{k=0}^{\infty} (I - A^{-1}B)^k \text{ 存在,且}$$

$$\sum_{k=0}^{\infty} (I - A^{-1}B)^k = (I - (I - A^{-1}B)^{-1}) = B^{-1}A,$$

所以
$$\sum_{k=0}^{\infty} (I - A^{-1}B)^k \cdot A^{-1} = B^{-1}A \cdot A^{-1} = B^{-1}.$$

其次证明 $\left\|\sum_{k=0}^{\infty} (I - A^{-1}B)^k \cdot A^{-1}\right\| \leqslant \dfrac{\alpha}{1 - \alpha\beta}$. (9 − 14)

因为
$$\left\|\sum_{k=0}^{\infty} (I - A^{-1}B)^k \cdot A^{-1}\right\| \leqslant \left\|\sum_{k=0}^{\infty} (I - A^{-1}B)^k\right\| \cdot \|A^{-1}\|,$$

而
$$\sum_{k=0}^{\infty} (I - A^{-1}B)^k = \sum_{k=0}^{\infty} [A^{-1}(A - B)]^k = \sum_{k=0}^{\infty} (A^{-1})^k \cdot (A - B)^k,$$

故
$$\left\|\sum_{k=0}^{\infty} (I - A^{-1}B)^k\right\| \leqslant \sum_{k=0}^{\infty} \|A^{-1}\|^k \cdot \|A - B\|^k \leqslant \sum_{k=0}^{\infty} (\alpha\beta)^k$$

$$=\lim_{k\to\infty}\frac{1-(\alpha\beta)^{k-1}}{1-\alpha\beta}=\frac{1}{1-\alpha\beta},$$

此即(9－14)成立．由(9－13)立即可知(9－12)成立．定理证毕．

定理 9.2　设

(1) $S(x)=f(x)^{T}f(x)$，x^{*}，为 $S(x)$ 的稳定点，其中，$f(x)$ 在 $\Omega=\{x\mid \|x-x^{*}\|<\delta$ 内有连续三阶偏导数；

(2) $F(x)=\nabla S(x)$ 的 Jacobi 矩阵 $J(x)$ 非奇异且对 $\forall x\in\Omega$ 有 $\|J(x)^{-1}\|_M<\omega$；

(3) $|\partial_i\partial_j g_k(x)|\leqslant v(i,j,k=1,\cdots,n)$　(对 $\forall x\in\Omega$)；

(4) $|\partial_i\partial_j f_k(x)|\leqslant\lambda(k=1,\cdots,m;i,j=1,\cdots,n)$ (对 $\forall x\in\Omega$)；

(5) $\|f(x)-f(x^{*})\|_M\leqslant L\|x-x^{*}\|_M$(对 $\forall x\in\Omega$)；

(6) $\mu\|x-x^{*}\|_M\leqslant r<\frac{1}{2}$，其中 $\mu=\max(2\omega mnL\lambda,\omega vn^{2})$；

(7) $f(x^{*})=0$，

则当 $x_0\in\Omega$，按 G-N 算法 9－1 所构成的序列 $\{x_k\}$ 收敛于 x^{*}，且有以下估计式：

$$\|x_{k+1}-x^{*}\|_M<2\mu\|x_k-x^{*}\|_M^2\quad(\forall k\geqslant 0).$$

本定理采用数学归纳法证明．在给出证明之前先对定理的条件及证明思路给予如下说明：

1. $J(x)$ 是 $\nabla S(x)$ 的 Jacobi 矩阵，亦即 $S(x)$ 的 Hesse 矩阵，即

$$J(x)=\nabla^{2}S(x)=2\left[A(x)^{T}A(x)+\left(\sum_{l=1}^{m}f_l(x)\nabla^{2}f_l(x)\right)\right].$$

令 $Q(x)=A(x)^{T}A(x)$，$D(x)=\left(\sum_{l=1}^{m}f_l(x)\nabla^{2}f_l(x)\right)$.

记
$$Q(x)=(q_{ij}(x))_{i,j=1,\cdots,n},$$

其中
$$q_{ij}(x)=\sum_{l=1}^{m}\partial_i f_l(x)\partial_j f_l(x).$$

记
$$D(x)=(d_{ij}(x))_{i,j=1,\cdots,n},$$

其中
$$d_{ij}(x)=\sum_{l=1}^{m}f_l(x)\partial_j\partial_i f_l(x).$$

给出 $J(x)$ 非奇异，且 $\|J(x)^{-1}\|_M<\omega$ 是为了导出 $Q(x)$ 非奇异，且 $\|Q(x)^{-1}\|\leqslant\frac{2\omega}{2(1-r)}$；

2. 条件(3)—(7)都是为了估计归纳结论中的不等式而设的 $f(x^*)=0$ 表示问题为零残量,因此本定理的二阶收敛速率的结论只是对零残量问题才是精确的;

3. 本定理欲证明从 $x_0\in\Omega$ 出发,按算法 7-1 所产生的点列 $x_k\to x^*(k\to\infty)$,即对 $\forall\varepsilon>0$,存在 K,当 $k>K$ 时有 $\|x_k-x^*\|_M<\varepsilon$ 且 $\|x_{k+1}-x^*\|_M<2\mu\|x_k-x^*\|_M^2(\forall k\geqslant 0)$.

上述结论用数学归纳法证明　即对 $\forall k$,若 $x_k\in\Omega$,有

$$\|x_{k+1}-x^*\|_M\leqslant\alpha\|x_k-x^*\|_M,\tag{9-15}$$

$$\|x_{k+1}-x^*\|_M\leqslant\frac{\mu}{1-r}\|x_k-x^*\|_M^2,\tag{9-16}$$

其中 $\alpha=\dfrac{r}{1-r}<1$.

由(9-15)可得

$$\|x_k-x^*\|_M\leqslant\alpha^k\|x_0-x^*\|_M\quad(\forall k\geqslant 0),$$

从而当 $k\to 0$ 时,$\|x_k-x^*\|_M\to 0$,即　$\lim\limits_{k\to\infty}x_k=x^*$,

再由(9-16),且注意到 $\dfrac{1}{1-r}<2$,有

$$\|x_{k+1}-x^*\|_M\leqslant 2\mu\|x_k-x^*\|_M^2\quad(\forall k\geqslant 0).$$

下面我们验证归纳法的第一步,即当 $x_0\in\Omega$ 时,有

$$\|x_1-x^*\|_M\leqslant\alpha\|x_0-x^*\|_M,\tag{9-17}$$

及

$$\|x_1-x^*\|_M\leqslant\frac{\mu}{1-r}\|x_0-x^*\|_M^2.\tag{9-18}$$

首先须证明 $Q(x_0)^{-1}$ 存在且有估计式

$$\|Q(x_0)^{-1}\|_M\leqslant\frac{2\omega}{2(1-r)}.\tag{9-19}$$

事实上,$J_0=2(Q_0+D_0)$,$D_0=\dfrac{1}{2}J_0-Q_0$,由条件 2 知,J_0 非奇异,

且

$$\left\|\frac{1}{2}J_0^{-1}\right\|_M\leqslant\frac{\omega}{2}.\tag{9-20}$$

又因

$$\left\|\frac{1}{2}J_0-Q_0\right\|_M=\|D_0\|_M=\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}\left|\sum_{l=1}^{m}f_l(x_0)\partial_j\partial_i f_l(x_0)\right|$$

$$\leqslant\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}\sum_{l=1}^{m}|f_l(x_0)|\cdot|\partial_j\partial_i f_l(x_0)|,\tag{9-21}$$

据(9-21)并利用条件 4,条件 5 及条件 7 进一步得

$$\left\|\frac{1}{2}J_0-Q_0\right\|_M\leqslant\lambda mnL\|x_0-x^*\|_M,\tag{9-22}$$

若取 $\mu = 2\omega mnL\lambda$，再用条件 6，则(9 - 22)变为

$$\left\|\frac{1}{2}J_0 - Q_0\right\|_M \leqslant \frac{r}{2\omega}, \tag{9-23}$$

在(9 - 20)及(9 - 23)中

$$\frac{\omega}{2} \cdot \frac{r}{2\omega} = \frac{r}{4} < 1,$$

故由定理 9.1 知，Q_0 非奇异，且

$$\|Q_0^{-1}\|_M \leqslant \frac{\frac{\omega}{2}}{1 - \frac{r}{2\omega} \cdot \frac{\omega}{2}} = \frac{2\omega}{4 - r} < \frac{2\omega}{2(2 - r)},$$

即(9 - 19)成立.

其次证明

$$\|x_1 - x^*\|_M \leqslant \frac{r}{i - r}\|x_0 - x^*\|_M, \tag{9-24}$$

事实上，由于

$$\begin{aligned} x_1 - x^* &= (x_1 - x_0) + (x_0 - x^*) = -Q_0^{-1}g_0 + (x_0 - x^*) \\ &= Q_0^{-1}(Q_0(x_0 - x^*) - g_0) \\ &= Q_0^{-1}(Q_0(x_0 - x^*) + (g^* - g_0)), \end{aligned}$$

故

$$\begin{aligned} \|x_1 - x^*\|_M &\leqslant \|Q_0^{-1}\|_M \cdot \|Q_0(x_0 - x^*) + (g^* - g_0)\|_M \\ &= \|Q_0^{-1}\|_M \cdot \left\|\frac{1}{2}(J_0 - D_0)(x_0 - x^*) + (g^* - g_0)\right\|_M \\ &\leqslant \frac{2\omega}{1 - r}\left(\left\|\frac{1}{2}J_0(x_0 - x^*) + (g^* - g_0)\right\|_M \right. \\ &\quad \left. + \|D_0(x_0 - x^*)\|\right), \end{aligned} \tag{9-25}$$

在(9 - 25)中，易证

$$\left\|\frac{1}{2}J_0(x_0 - x^*) + (g^* - g_0)\right\|_M \leqslant \frac{vM^2}{2}\|x_0 - x^*\|_M^2,$$

又

$$\|D_0\|_M \leqslant \lambda mnL\|x_0 - x^*\|_M,$$

于是得

$$\begin{aligned} \|x_1 - x^*\|_M &\leqslant \frac{2\omega}{2(1 - r)}\left(\frac{vn^2}{2} + \lambda mnL\right) \cdot \|x_0 - x^*\|_M^2 \\ &\leqslant \frac{1}{2(1 - r)}(\omega vn^2 + 2\omega\lambda mnL)\|x_0 - x^*\|_M^2 \\ &\leqslant \frac{\mu}{1 - r}\|x_0 - x^*\|_M^2 \leqslant \frac{r}{1 - r}\|x_0 - x^*\|_M^2. \end{aligned}$$

取 $\alpha = \frac{r}{1-r} < 1$，于是(9－17)及(9－18)同时成立.

仿照上述证明过程，读者不难证明(9－15)及(9－16)成立.[注]

例 9.3 设 $S(x) = (x+1)^2 + (\lambda x^2 + x - 1)^2, x \in \mathbf{R}^1$，试讨论函数的极小点.

解 这是一个一元问题($n=1, m=2$)，令 $f_1(x) = x+1, f_2(x) = \lambda x^2 + x - 1$，其中 x 是问题所要确定的参数，而 λ 是任意常数.

用古典分析的方法容易得到，当 $\lambda < 1$ 时，$x = 0$ 为 $S(x)$ 的一个极小点. 现在用 G-N 方法讨论 $S(x)$ 的极小点.

设初始点为 x_k，计算：

$$A_k = Df(x_k) = \left(\frac{\partial f_1}{\partial x}, \frac{\partial f_2}{\partial x}\right)^{\mathrm{T}}_{x=x_k} = (1, 2\lambda x_k + 1)^{\mathrm{T}};$$

$$A_k^{\mathrm{T}} A_k = 1 + (2\lambda x_k + 1)^2;$$

$$A_k^{\mathrm{T}} f_k = 2\lambda^2 x_k^3 + 3\lambda x_k^2 - 2\lambda x_k + 2x_k.$$

由 $A_k^{\mathrm{T}} A_k \Delta x_k = -A_k^{\mathrm{T}} f_k$，得

$$\Delta x_k = -\frac{2\lambda^2 x_k^3 + 3\lambda x_k^2 - 2\lambda x_k + 2x_k}{1 + (2\lambda x_k + 1)^2},$$

所以

$$x_{k+1} = x_k + \Delta x_k = \frac{2\lambda^2 x_k^3 + \lambda x_k^2 + 2\lambda x_k}{1 + (2\lambda x_k + 1)^2}. \tag{9－26}$$

(1) 当 $\lambda = 0$ 时，对于任给的初始点 x_k，有 $x_{k+1} = 0$，即是说，用 G-N 方法经过一次迭代便得到极小点 $x_{k+1} = x^* = 0$.

(2) 当 $\lambda \neq 0$，且 $|\lambda| < 1$ 时，若假定初始近似 x_k 充分小，则分别将(9－26)的分子分母写成：

$$2\lambda x_k + \lambda x_k^2 + 2\lambda^2 x_k^3 = 2\lambda x_k + o(x_k),$$

$$1 + (2\lambda x_k + 1)^2 = 2 + o(x_k),$$

故

$$x_{k+1} = \lambda x_k + o(x_k^2). \tag{9－27}$$

由(9－27)反复迭代有

$$x_{k+p} = \lambda^p x_k + o(x_k^2),$$

因为 $|\lambda| < 1$，故当 $p \to \infty$ 时，有

$$x_{k+p} \to 0,$$

即是说，当 $\lambda \neq 0$，且 $|\lambda| < 1$ 时，若取初始近似 x_k 充分小，则由 G-N 迭代式(9－27)所产生的点列 $\{x_k\}$ 收敛于 $x^* = 0$，而且容易验证 $\{x_k\}$ 线性收敛于 $x^* = 0$.

注：上述定理中所用范数 $\|\cdot\|_W$，亦即 $\|\cdot\|_\infty$.

(3) 当 $\lambda < -1$ 时,由古典分析方法可以得到 $S(\boldsymbol{x})$ 的极小点 $\boldsymbol{x}^*$ 仍为 0,但由 G-N 迭代式(9-27)却不能得到 $\boldsymbol{x}^* = 0$,因为此时 $\{\boldsymbol{x}_k\}$ 不收敛.

9.3　修正的 G-N 算法(Hartley 方法)

如前所述,G-N 算法是一个局部收敛方法.它对初始点 $\boldsymbol{x}_0$ 的依赖性很大,即仅当 $\boldsymbol{x}_0$ 充分接近极小点 $\boldsymbol{x}^*$ 时才有可能收敛.修正方案有二:

1. 采用完备算法

给定初始点 $\boldsymbol{x}_0$,令 $k=0,1,\cdots$,计算

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k, \tag{9-28}$$

其中 $\boldsymbol{p}_k = -(\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k)^{-1}\boldsymbol{g}_k$,$\alpha_k$ 满足

$$S(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) = \min_{\alpha} S(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k).$$

在(9-28)中,由于 α_k 由精确的一维寻查确定,从而使算法放宽对初值 $\boldsymbol{x}_0$ 的要求,起到了扩大收敛域的作用,但在每一迭代步都进行一维寻查,其计算量是相当大的.

2. 采用可接受点算法

给定初始点 $\boldsymbol{x}_0$,令 $k=0,1,\cdots$,计算

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k, \tag{9-29}$$

其中 $\boldsymbol{p}_k = -(\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k)^{-1}\boldsymbol{g}_k$,$\alpha_k$ 满足

$$S(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) < S(\boldsymbol{x}_k),$$

由于 $\boldsymbol{p}_k \neq 0$ 为 $S(\boldsymbol{x})$ 处的下降方向,据下降方向的性质,对于小正数 $\delta > 0$,存在 $\alpha_k \in (0,\delta)$ 满足

$$S(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) < S(\boldsymbol{x}_k), \tag{9-30}$$

但直接从(9-30)所定出的 α_k 很小,从而增大了计算量.

1972 年,Meyer 提出用另外的不等式替代(9-30),有如下定理:

定理 9.3　设 $S(\boldsymbol{x})$ 为可微的严格凸函数,则对于 $\beta \in (0,1)$ 的每一个值,总存在一个区间 $(0,\hat{\alpha})$ 使下式成立:

$$S(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) \leqslant S(\boldsymbol{x}_k) + \beta\alpha \boldsymbol{p}_k^{\mathrm{T}} \nabla S(\boldsymbol{x}_k) = S(\boldsymbol{x}_k) + 2\beta\alpha \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{g}_k, \tag{9-31}$$

其中 $\boldsymbol{p}_k = \Delta \boldsymbol{x}_k$ 为下降方向.

上述定理的一般证明从略,下面就 $S(\boldsymbol{x})$ 为凸二次函数的情形给予几何上的证明.

令 $S(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) = \varphi(\alpha)$,于是(9-31)变形为

$$\varphi(\alpha) \leqslant \varphi(0) + 2\beta\alpha\, \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{g}_k, \tag{9-32}$$

当 $\beta=1$ 时，成立等式

$$\varphi(\alpha) \leqslant \varphi(0) + 2\alpha p_k^{\mathrm{T}} g_k,$$

即
$$\varphi(\alpha) \leqslant \varphi(0) + \varphi'(0)\alpha. \tag{9-33}$$

(9－33)是抛物线 $\varphi(\alpha)$ 在 $\alpha=0$ 处的切线 TU(图 9－2)的方程.

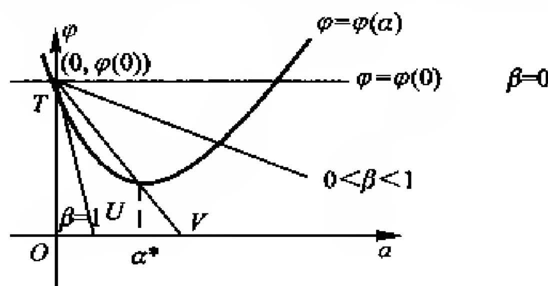


图 9－2

当 $\beta=0$ 时，$\varphi(\alpha)=\varphi(0)$，此时 $\varphi(\alpha)$ 退化成水平线. 因此我们只考虑当 $\beta\in(0,1)$ 时，寻找 $\alpha\in(0,\hat{\alpha})$，使(9－32)亦即(9－31)成立.

例如，容易证明，对于凸二次函数 $\varphi(\alpha)$，若取 $\beta=\frac{1}{2}$，则存在 $[0,\alpha^*]$ (其中 α^* 为 $\varphi(\alpha)$ 的极小点)，当 $\alpha\in[0,\alpha^*]$ 时(9－32)成立.

当 $\beta=\frac{1}{2}$ 时，在(9－32)中取等号，则有

$$\varphi(\alpha) = \varphi(0) + \alpha p_k^{\mathrm{T}} g_k = \varphi(0) + \frac{1}{2}\alpha\varphi'(0). \tag{9-34}$$

(9－34)即是二次曲线 $\varphi(\alpha)$ 过 $(0,\varphi(0))$ 及 $(\alpha^*,\varphi(\alpha^*))$ 两点的割线.

从图(9－2)看出，当 β 从 1 减至 0 时，区间 $[0,\hat{\alpha}]$ 的范围随之变大.

下面进一步解释不等式(9－31)的作用，在(9－31)中，因为 p_k 为下降方向，则由于 $p_k^{\mathrm{T}} g_k<0$，α、$\beta>0$，故式中第二项为负，即

$$S(x_k+\alpha_k p_k) < S(x_k),$$

并且由于 α、$\beta>0$ 为适当小的正数，致使 $S(x_k+\alpha_k p_k)-S(x_k)$ 保持一定的数值，即是说满足(9－31)的 α_k 不仅保证 $S(x)$ 过 x_k 沿 $\alpha_k p_k$ 方向移动时函数值下降，而且使下降量保证一定的数量，这就使在保证函数值下降的前提下 $\alpha_k p_k$ 不会太小，从而减少迭代的次数.

1970 年，Meyer 还进一步证明，α_k 满足(9－31)的可接受点算法所产生点列 $\{x_k\}$ 是收敛的. 上述算法由于避免了精确的一维寻查而受到重视.

算法 9.2(Hartley-Meyer 算法)

只将算法 9.1 的第六步如下修改，其余步骤不变：

令
$$x_{k+1}=x_k+\alpha_k p_k,$$

其中 α_k 通过如下步骤确定：

① 令 $\alpha_k=1,\beta=10^{-5}$；

② 检查 $S(\boldsymbol{x}_k+\alpha\boldsymbol{p}_k)\leqslant S(\boldsymbol{x}_k)+2\beta\alpha\boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{g}_k$，

是，令 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k$，转算法 9.1 的 ⑦；

③ 置 $\alpha_k=\dfrac{\alpha_k}{2}$，转②.

9.4 Levenberg-Marquarat 算法(简称 L-M 算法)

修正 G-N 算法的另一条途径是改善矩阵 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 的性态. 我们知道在 G-N 方向

$$\boldsymbol{p}_k=-(\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k)^{-1}\boldsymbol{g}_k$$

中，$\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 的性态对 $\boldsymbol{p}_k$ 的影响很大，当 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 奇异时，$\boldsymbol{p}_k$ 的构造遭到破坏，从而使迭代进行不下去；当 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 病态时，通过法方程

$$\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k\boldsymbol{p}_k=-\boldsymbol{g}_k$$

所得的解会严重失真.

为了克服上述两种异常情形，有以下两种方案可供选择：

1. 在每次迭代中，均用 Golub 方法解线性最小二乘问题 $\min\|\boldsymbol{A}_k\boldsymbol{p}_k+f_k\|_2^2$. 这样做的缺点是增加了计算复杂性，尤其是 $\boldsymbol{A}_k$ 列降秩，即 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 奇异时尤为突出；其次采用 Golub 算法失去了灵活性，即对于某些迭代步，尽管 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 条件较好，但仍当成病态处理.

2. 采用对系数矩阵阻尼的方法，这种做法不是放弃法方程而是改造法方程，使其条件变好，并且对于某些迭代步，当其 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 条件较好时，即恢复为 G-N 算法.

9.4.1 L-M 算法的基本想法

L-M 算法的基本想法是，通过对 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 的正则化改善其条件. 容易验证，对于正半定矩阵 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$，当 $\mu>0$ 时，$\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k+\mu\boldsymbol{I}$ 为正定，因此，我们可以选择适当大的 $\mu>0$，修改法方程为

$$(\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k+\mu\boldsymbol{I})\boldsymbol{p}_k=-\boldsymbol{A}_k^{\mathrm{T}}f_k, \tag{9-35}$$

称(9－35)的解，即

$$\boldsymbol{p}_k(\mu)=-(\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k+\mu\boldsymbol{I})^{-1}\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k \tag{9-36}$$

为修小的 G-N 方向. 往后将证明，$\boldsymbol{p}_k(\mu)$ 的模随 μ 的增大而减少. 因此取适当大的 $\mu>0$，以(9－36)为寻查方向，$S(\boldsymbol{x})$ 的值将减少，即是说如下算

法的下降性得到保证：

$$x_k \text{ 给定}, x_{k+1} = x_k + p_k(\mu)$$

$$p_k(\mu_k) = -(A_k^T A_k + \mu I)^{-1} A_k^T f_k, \quad k = 0,1,\cdots. \tag{9-37}$$

(9-37)即是 L-M 算法的粗略轮廓. 显然该算法的关键在于每次迭代时，如何选择适当的 $\mu_k > 0$ 以保证其下降性质. 为此，需要弄清楚如下几个问题：

1. 修正的 G-N 方向 $p_k(\mu)$ 与线性最小二乘问题 $\min \|A_k p_k + f_k\|_2^2$ 有什么关系?

2. $p_k(\mu)$ 与 $\mu > 0$ 的关系?

3. $p_k(\mu)$ 与最速下降方向 $-g_k = -A_k^T f_k$ 的关系?

4. $S(x)$ 过 x_k 沿 $p_k(\mu)$ 移动时的下降量?

解决上述四个问题构成了 L-M 算法的全部理论.

9.4.2 L-M 算法的基础定理

定理 9.4 设(1) $\mu \geqslant 0$ 为给定的实数；

(2) $p_k(\mu)$ 为方程 $(A_k^T A_k + \mu I) p_k = -A_k^T f_k$ 的解，其中 $p_k = x - x_k$，则 $p_k(\mu)$ 为二次函数 $S(p_k) = \|A_k p_k + f_k\|_2^2$，在闭球 $\|p_k\|_2^2 \leqslant \|p_k(\mu)\|_2^2$ 上的最优解，即

$$\hat{S}(p_k(\mu)) = \min_{p_k \in \bar{B}_k} \hat{S}(p_k),$$

其中

$$\bar{B}_k = \bar{B}_k(x_k, \|p_k(\mu)\|_2).$$

证：$\hat{S}(p_k) = \|A_k p_k + f_k\|_2^2 = f_k^T f_k + 2(A_k^T f_k)^T + p_k^T A_k^T A_k p_k$，

由条件(2)，得

$$\hat{S}(p_k) = f_k^T f_k - 2\mu p_k(\mu)^T p_k - 2 p_k(\mu)^T A_k^T A_k p_k + p_k^T A_k^T A_k p_k,$$

而
$$\hat{S}(p_k(\mu)) = f_k^T f_k - 2\mu p_k(\mu)^T p_k(\mu) - 2 p_k(\mu)^T A_k^T A_k p_k(\mu),$$

故 $\hat{S}(p_k) - \hat{S}(p_k(\mu))$

$$= (p_k(\mu) - p_k)^T A_k^T A_k (p_k(\mu) - p_k) + \mu(p_k(\mu) - p_k)^T \cdot (p_k(\mu) - p_k)$$
$$+ \mu(p_k(\mu)^T p_k(\mu) - p_k^T p_k).$$

因为 $A_k^T A_k$ 正半定，故对于 $\forall p_k \in \bar{B}_k$，由于 $p_k(\mu) - p_k \neq 0$，有

$$(p_k(\mu) - p_k)^T A_k^T A_k (p_k(\mu) - p_k) \geqslant 0,$$

及
$$(p_k(\mu) - p_k)^T (p_k(\mu) - p_k) > 0,$$

又因 $p_k \in \bar{B}_k$，有

$$p_k^T(\mu) p_k(\mu) - p_k^T p_k = \|p_k(\mu)\|_2^2 - \|p_k\|_2^2 \geqslant 0,$$

所以对于 $\forall p_k \in \overline{B}_k$，有

$$\hat{S}(p_k) - \hat{S}(p_k(\mu)) \geqslant 0,$$

即 $p_k(\mu)$ 为 $\hat{S}(p_k)$ 在 $\hat{B}_k$ 上的极小点.

定理 9.4 表明，对法方程 $A_k^T A_k p_k = -A_k^T f_k$ 加入阻尼项，相当于在原有线性最小二乘问题上加上约束条件，其作用在于，当 $A_k^T A_k$ 奇异，因而 $A_k^T A_k p_k = -A_k^T f_k$ 无确定解，即 $\hat{S}(p_k)$ 在 $\mathbf{R}^n$ 上无唯一极小点时，通过阻尼步骤，使 $\hat{S}(p_k)$ 在 B_k 上获得唯一极小点 $p_k(\mu)$，尽管 $\hat{S}(p_k)$ 不能在 p_k 上获得最小值(最大限度的下降)，但能在 p_k 方向上获得约束条件下的极小值，从而达到下降的目的.

定理 9.5 设(1) $\mu>0$ 为给定的实数；

(2) $p_k(\mu)$ 为方程 $(A_k^T A_k+\mu I)p_k = -A_k^T f_k$ 的解.

则 $\|p_k(\mu)\|_2^2$ 是 μ 的连续下降函数，且当 $\mu\to\infty$ 时，$\|p_k(\mu)\|_2^2\to 0$.

证 $A_k^T A_k$ 为实对称半正定矩阵，故存在正交阵 C，使得

$$C^T A_k^T A_k C = M = \mathrm{diag}(\lambda_1,\cdots,\lambda_n),$$

其中 $\lambda_i \geqslant 0(i=1,\cdots,n)$ 为 $A_k^T A_k$ 的特征值.

由条件 2
$$(A_k^T A_k+\mu I)p_k = -A_k^T f_k,$$
得
$$C^T(A_k^T A_k+\mu I)CC^T p_k = -C^T A_k^T f_k,$$
即
$$(M+\mu I)C^T p_k = -C^T A_k^T f_k.$$

因为 $\mu>0$，且 M 为半正定，所以 $M+\mu I$ 为正定，从而 $(M+\mu I)^{-1}$ 存在，故有

$$p_k(\mu) = -C(M+\mu I)^{-1}C^T A_k^T f_k.$$

令
$$w = C^T A_k^T f_k,$$
则
$$p_k(\mu) = -C(M+\mu I)^{-1}w,$$

所以
$$\begin{aligned}\|p_k(\mu)\|_2^2 &= p_k(\mu)^T p_k(\mu) = w^T(M+\mu I)^{-1}C^T C(M+\mu I)^{-1}w \\ &= w^T((M+\mu I)^{-1})^2 w = w^T((M+\mu I)^2)^{-1}w \\ &= \sum_{i=1}^n \frac{w_i^2}{(\mu+\lambda_i)^2},\end{aligned}$$

其中 $w=(w_1,\cdots,w_n)^T$.

由于 $w_i,\lambda_i \geqslant 0(i=1,\cdots,n)$ 均为与 μ 无关的常数，故

$$\|p_k(\mu)\|_2^2 = \sum_{i=1}^n \frac{w_i^2}{(\mu+\lambda_i)^2}$$

是 μ 的连续单调下降函数，且当 $\mu\to\infty$ 时，$\|p_k(\mu)\|_2^2\to 0$.

容易验证　　$p_{i1}\cdot p_{i3}=p_{i2}^2.$

现将(9－39)的分子的三个式子换成

$$\sum_{i=1}^{n}\frac{w_i^2}{(\mu+\lambda_i)}=\frac{\sum_{i=1}^{n}w_i^2\prod_{j\neq i}(\mu+\lambda_j)}{\prod_{j=1}^{n}(\mu+\lambda_j)}=\frac{\sum_{i=1}^{n}w_i^2p_{i1}}{\prod_{j=1}^{n}(\mu+\lambda_j)},$$

$$\sum_{i=1}^{n}\frac{w_i^2}{(\mu+\lambda_i)^2}=\frac{\sum_{i=1}^{n}w_i^2\prod_{j\neq i}(\mu+\lambda_j)^3}{\prod_{j=1}^{n}(\mu+\lambda_j)^3}=\frac{\sum_{i=1}^{n}w_i^2p_{i3}}{\prod_{j=1}^{n}(\mu+\lambda_j)^3},$$

$$\sum_{i=1}^{n}\frac{w_i^2}{(\mu+\lambda_i)^2}=\frac{\sum_{i=1}^{n}w_i^2p_{i2}}{\prod_{j=1}^{n}(\mu+\lambda_j)^2},$$

所以

$$\phi'(\mu)=\frac{\left(\sum_{i=1}^{n}w_i^2p_{i1}\right)\left(\sum_{i=1}^{n}w_i^2p_{i3}\right)-\left(\sum_{i=1}^{n}w_i^2p_{i2}\right)^2}{\left[\sum_{i=1}^{n}\frac{w_i^2}{(\mu+\lambda_i)^2}\right]^{\frac{3}{2}}\left[\sum_{i=1}^{n}(\mu+\lambda_i)^2\right]^2\left(\sum_{i=1}^{n}w_i^2\right)^{\frac{1}{2}}},\tag{9－41}$$

在(9－41)中，对于 $\mu>0$，分母是有意义的，且为正数，故 $\phi'(\mu)$ 的符号取决于分子的符号. 利用 Cauchy-Schwars 不等式，可以证明

$$\left(\sum_{i=1}^{n}w_i^2p_{i1}\right)\left(\sum_{i=1}^{n}w_i^2p_{i3}\right)-\left(\sum_{i=1}^{n}w_i^2p_{i2}\right)^2\geqslant 0,$$

因此，对于 $\mu>0$，$\phi'(\mu)\geqslant 0$，即 $\phi(\mu)$ 为 μ 的连续单调上升函数，且由(9－38)知，当 $\mu>0$ 时

$$\phi(\mu)=\frac{\sum_{i=1}^{n}\frac{w_i^2}{\left(1+\frac{\lambda_i}{\mu}\right)}}{\left[\sum_{i=1}^{n}\frac{w_i^2}{\left(1+\frac{\lambda_i}{\mu}\right)}\right]^{\frac{1}{2}}\left(\sum_{i=1}^{n}w_i^2\right)^{\frac{1}{2}}},$$

故当 $\mu\to\infty$ 时，$\phi(\mu)\to 1$.

定理 9.5 及定理 9.6 表明，修正的 G-N 方向，即 $\boldsymbol{p}_k(\mu)$ 当 $\mu>0$ 时，其大小 $\|\boldsymbol{p}_k(\mu)\|_2$ 随 $\mu\to\infty$ 而趋于 0；其方向随 $\mu\to\infty$ 而与 $\boldsymbol{g}_k=-\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{f}_k=-\frac{1}{2}\nabla S(\boldsymbol{x}_k)$ 的方向一致($\phi(\mu)\to 1$，即 $(\boldsymbol{p}_k(\mu),\boldsymbol{g}_k)\to 0$). 如图 9－3 所示.

(2) $Q(x)=A(x)^{T}A(x)$ 在 $C(x_0)$ 上正定;

(3) $\{x_k\}$ 是以 x_0 为初始点的 L-M 序列,

则 $\{x_k\}$ 必有极限点,且它的所有极限点都是 $S(x)$ 的稳定点.

证　首先证明 $\{x_k\}$ 必存在极限点.

因为 $\{x_k\}$ 为 L-M 序列,故对 $\forall k$,有 $S(x_{k+1})<S(x_k)$,

所以
$$\{x_k\}\subset C(x_0).$$

而据条件(1),$\{x_k\}$ 中必存在收敛的子序列 $\{x_{k_i}\}$,即当 $k_i\to\infty$ 时,
$$x_{k_i}\to x^*\subset C(x_0),$$
因此 $\{x_k\}$ 在 $C(x_0)$ 中必存在极限点.

其次,设 x^* 为 $\{x_k\}$ 的任一极限点,用反证法证明 x^* 为 $S(x)$ 的驻点.

设相反,x^* 非 $S(x)$ 的驻点,即 $g(x^*)\neq 0$.

由 $Q(x^*)$ 正定,$Q(x^*)+\mu I(\mu>0)$ 亦正定有
$$g^{*T}p^*=-g^{*T}(Q(x^*)+\mu I)^{-1}g^*<0. \tag{9-43}$$
另一方面,因 $S(x)$ 满足定理 9.7 的条件,故对于任意给定 $\beta\in(0,1)$,总存在 $\mu>0$,使
$$S(x_k+p_k(\mu))\leqslant S(x_k)+2\beta p_k(\mu)^{T}g_k. \tag{9-44}$$

将 x_{k_i} 代入(9-43),有
$$S(x_{k_i}+p_{k_i})\leqslant S(x_{k_i})+2\beta p_{k_i}^{T}g_{k_i}. \tag{9-45}$$

因为
$$0\leqslant S(x_{k+1})<S(x_k),$$
即 $\{S(x_k)\}$ 为单调下降有界序列,故必有极限,

所以
$$\lim_{k_i\to\infty}S(x_{k_i+1})=\lim_{k_i\to\infty}S(x_{k_i}),$$
又由 $S(x)$ 的连续有
$$\lim_{k_i\to\infty}S(x_{k_i+1})=\lim_{k_i\to\infty}S(x_{k_i})=S(x^*),$$
再考虑到 $g(x)$ 的连续性,在(9-45)中令 $k_i\to\infty$,有
$$S(x^*)\leqslant S(x^*)+2\beta p^{*T}g^*,$$
即
$$2\beta p^{*T}g^*\geqslant 0,$$
亦即
$$p^{*T}g^*\geqslant 0, \tag{9-46}$$
由于(9-46)与(9-43)矛盾,故得
$$g(x^*)=0.$$

定理 9.8 只表明了 $\{x_k\}$ 的极限点为 $S(x)$ 的稳定点，如果再加强条件，便可证得 $\{x_k\}$ 收敛于 $S(x)$ 的稳定点.

定理 9.9　设(1) $S(x)$ 对于 x_0 的水平集 $C(x_0)=\{x \mid S(x)\leqslant S(x_0)\}\subset \mathbf{R}^n$ 是列紧的；

(2) $S(x)$ 在 $C(x_0)$ 上至多有有限个稳定点，使之在其上有任意给定的值；

(3) $Q(x)$ 在 $C(x_0)$ 上正定；

(4) $\{x_k\}$ 是以 x_0 为初始点的 L-M 序列，

则 $\{x_k\}$ 收敛于 $S(x)$ 的稳定点.

证明略，请参阅参考文献[1]第七章定理 7.7.2.

定理 9.10　设(1) $\{x_k\}$ 为 L-M 序列；

(2) $\{x_k\}$ 收敛于 $S(x)$ 的某个稳定点 x^*；

(3) q 为 $Q(x^*)=A(x^*)^{\mathrm{T}}A(x^*)$ 的最小特征值；

(4) $\gamma=\max\limits_{1\leqslant i\leqslant n}\{|\gamma_i|\}$，其中 $\gamma_i(i=1,\cdots,n)$ 为 n 阶矩阵 $R(x^*)$ 的特征值，这里 $R(x)=\sum\limits_{i=1}^{n} f_i(x)\ \nabla^2 f_i(x)$；

(5) $$\mu<\frac{r}{q}<1;$$

(6) $$\beta<\frac{1-\mu}{2},$$

则有(1) x^* 为 $S(x)$ 的严格极小点；

(2) $$\lim_{k\to\infty}\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}\leqslant\mu.$$

证明略，请参阅参考文献[1]第七章定理 7.7.3.

附录Ⅱ 最优化方法的发展进程

最优化方法和理论来源于军事、管理和经济. 我国古代军事天才孙武所著《孙子兵法》成书于公元前5世纪春秋末期，是我国也是世界上最古老的军事理论著作.《孙子兵法》十三篇把军事运筹学的思想、理论和方法阐述得淋漓尽致，可谓达到了神化的地步，连外国人也对孙武子倍加赞赏，1984年美国军事运筹学会出版专著"Systems Analysis and Modeling in Defence"，书中称孙武子是世界上第一个军事运筹学的实践家.

两次世界大战，尤其是二战中提出了很多军事最优化问题，诸如搜索潜艇问题、护航问题、布雷问题、轰炸问题以及运输问题等等. 二战中所提出的军事问题及其解决方法有如下特点：

数据是实践中的真实数据；解决问题的人员组成是多学科的；处理问题的方法渗透着物理学的思想.

第二次世界大战以后最优化方法的应用由军事问题转人民用问题，提出了现代管理的理论和方法，如计划管理、运输管理、工程建设管理、库存管理、工业工程管理等等.

1930年，苏联数学家康托诺维奇(Cantolovch)首创图上作业法解决小规模(以二维为主)的线性规划问题，如物资调运方案、合理下料方案等. 而大规模的线性规划(高维问题)则是在电子计算机问世以后才得以发展；

1936年，美国经济学家列昂节夫首创投入产出分析法，而反映国民经济动态运行的大型投人产出表的制定也是在大型计算机的支持下才能够实现；

1935—1940年，道奇(Dodge)和白莱盖(Blaket)首创统计抽样法，为此后数理统计学科发展奠定基础.

二战以后，科技和工业的蓬勃发展引起了许多新的变化，主要表现在：生产规模日益庞大、生产技术日益复杂、商业经营日趋国际化、环境污染和生态问题日益严重. 管理人员和计划人员面临十分复杂的问题，迫切需要一种统筹兼顾，行之有效的科学方法来处理复杂的管理问题. 最典型的例子即是邮递员问题(或者旅行商问题)及阿波罗登月发射工程.

例1 邮递员(旅行商问题)

一部邮车从一个城市向另外十个城市运送邮件，每一个城市只去一

次，可行路线为 10！ =3 628 800 条，这类问题理论上的可行路线上百万、上千万，如何寻求最短路径是实践中亟待解决的问题.动态规划方法为求解此类问题提供了数学手段.

例 2 阿波罗登月发射工程

由 300 多万个零件组成，动员 2 万多家企业，120 所大学和研究所参加，平均每年占用 42 万科技人员，历时 12 年（1960 年至 1972 年），耗资 250 亿美元，如此庞大的工程用传统的管理模式是无法想像的.

解决此类问题的数学模型即 PERT（计划评审技术），通过关键工序形成工程网络图确定优化的决策方案.

在 PERT 基础上发展起来的实用管理技术有二三十种，诸如 PERT/成本，PERT/可靠性，LCES（最低成本和估算计划），MAP（人力分配），PACP（产品分析控制）等等.国际上称赞此种管理技术为“经营管理科学上的决策性跃进”，“在管理方面的重要突破”.我国从 20 世纪 60 年代起，在电子计算机的研制、飞机总装配等方面采取此种管理技术取得了很大的成功.

下面着重介绍数学规划的一些重要分支的进展.

1. 线性规划（常用算法见本书第 10 ~ 11 章）

它是最优化方法发展最迅速，成就最大的一个分支.线性规划发展的爆炸时期是 20 世纪 50 年代至 60 年代，其奠基人应是前苏联数学家康托诺维奇（Cantolovch）和美国数学家丹茨（G. B. Dantzig）.

1947 年，Dantzig 提出了轰动数学界的单纯形法，为求解高维线性规划问题提供了一般的有效的工具；

1950—1965 年，匈牙利数学家库恩（H. W. Kuhn）和塔克（A. W. Tucker）建立了线性规划的对偶理论，为求解鞍点问题提供了数学工具，两位年轻的数学家建立约束极值的最优性条件被称为 K – T 条件，为求解非线性规划奠定了理论基础；

1958 年，美国数学家 R. E. Gomery 提出整数规划的割平面法；

1960 年，Dantzig 和英国数学家维尔夫（P. Wolfe）研究成功线性规划的分解算法，该算法为求解大规模线性规划提供了有力的工具；

1970—1984 年，前苏联数学家卡钦扬（L. G. Khachiyan）和美国数学家卡马尔卡（N. A. Karmarka）先后提出并完成了线性规划的多项式算法，轰动了整个数学界.

2. 非线性规划（20 世纪 70 年代的丰硕成果）

（1）无约束极值问题（常用算法见本书第 3 ~ 9 章）

1959—1963 年，由三位数学家共同研究成功求解无约束问题的 DFP

变尺度方法(该算法是由英国数学家 W. C. Davidon 提出,由法国数学家 R. Fletcher 和美国数学家 M. J. D. Powell 加以简化). 该算法的研究成功是无约束优化算法的一个大飞跃,引起了一系列的理论工作,并陆续出现了多种新的算法;

1965 年,德国数学家 C. G. Broyden 提出了求解非线性方程组的拟 Newton 法,并且该算法还包容了 DFP 算法;

1970 年,四位数学家从不同角度对变尺度算法进行深入研究,提出了 BFGS[注]公式. 实践表明该算法较之 DFP 算法和拟 Newton 法具有更好的数值稳定性.

1970 年,无约束优化方法的研究出现了引人注目的成果. 英国数学家 L. C. W. Dixon 和美籍华人 H. Y. Huang 提出了关于"二阶收敛算法的统一研究"的研究成果,该成果提出了一个含三个自由参数的公式族——Huang 族和拟 Newton 公式,它可包容前面所介绍的所有无约束优化算法.

例 3　香蕉函数(Banana Function)

$$f(\boldsymbol{x}) = f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2,$$

$$\boldsymbol{x}^{(0)} = (-1.9 \quad 2)^{\mathrm{T}}, \boldsymbol{x}^* = (1,1)^{\mathrm{T}}, f(\boldsymbol{x}^*) = 0.$$

在本书附录Ⅲ Matlab 的演示程序中,列出了求该二元函数极小点 $\boldsymbol{x}^*$ 的六种无约束算法:

DFP, BFGS, Steepest, Simplex, Gauss - Newton (G - N), Levenberg - Marquardt(L - M).

上述六种算法的效果比较:

从 $\boldsymbol{x}^{(0)}$ 出发能全程达到 $\boldsymbol{x}^*$ 的算法:DFP, BFGS, Simplex, L - M;

从 $\boldsymbol{x}^{(0)}$ 出发只能搜索到半途的算法:Steepest;

只能从 $\boldsymbol{x}^*$ 的邻点出发而达到 $\boldsymbol{x}^*$ 的算法:G-N;

从 $\boldsymbol{x}^{(0)}$ 出发,最快速达到 $\boldsymbol{x}^*$ 的算法:DFP, BFGS;

(2) 约束极值问题(常用算法见本书第 14 ~ 17 章)

约束极值的数学模型是

$$\min_{\boldsymbol{x} \in \Omega \subset \mathbf{R}^n} f(\boldsymbol{x}), \qquad f: \boldsymbol{x} \in \Omega \to \mathbf{R}^1$$

$$\Omega = \{\boldsymbol{x} \mid g_i(\boldsymbol{x}) \geqslant 0 \quad i = 1, 2, \cdots, m; h_j(\boldsymbol{x}) = 0, j = 1, 2, \cdots, l\}.$$

注: B——C. G. Broyden 德国数学家,
F——R. Fletcher 法国数学家,
G——D. Goldfarb 德国数学家,
S——D. F. Shanno 瑞士数学家

求解上述问题的基本算法是:罚函数法、可行方向法、投影梯度法.

20 世纪 70 年代以来,非线性规划算法有较大的进展,主要表现在三个方面:

第一,对罚函数方法的改进——乘子法和精确罚函数法

1970—1976 年,德国数学家 Hestenes 和美国数学家 Powell,Rockafellar 等提出了 Lagrange 乘子法及增广 Lagrange 乘子法,该种算法有效地克服了罚因子增大导致罚问题出现病态的缺陷.

为了克服罚函数法需要求解一系列无约束极小问题的缺陷,有人试图通过解单个无约束极值问题来代替原问题,即是找出一个精确罚函数.

第一个精确罚函数是荷兰数学家 W. I. Zangwill 提出来的,最初他是针对凸规划提出来的,以后又将此方法推广到非凸规划,但推广的算法遇到了麻烦,即在可行域的边界上往往是不可微的.为此法国数学家 Fletcher 首先研究了可微的精确罚函数算法. Fletcher 的工作把罚函数的改进算法推上了一个新的平台.

第二,对可行方向法的改进——"反锯齿措施"

1960 年,荷兰数学家 G. Zoutendijk 提出了线性约束下的可行方向法,但若将此法推广至非线性约束时则会发生所谓"锯齿现象",即算法收敛很慢,甚至收敛不到 K-T 点.

例 4 (Wolfe 反例)

20 世纪 60 年代,英国数学家 M. A. Wolfe 构造了一个具有唯一极小点的凸规划问题,但用 Zoutendijk 的方法计算出现了反常——锯齿现象.

$$(\mathrm{CP})\begin{cases}\min \dfrac{4}{3}(x_1^2 - x_1x_2 + x_2^2)^{\frac{3}{4}} + x_3,\\ \text{s.t } x_i \geqslant 0, i = 1,2,3.\end{cases}$$

该问题有唯一的 K-T 点 $x^* = (0,0,0)^T$,由于 x_i 的非负性,故 x^* 为问题的极小点.

若取初始点 $x^{(0)} = (a,0,c)^T$,其中 $0 \leqslant a \leqslant \frac{1}{4}$, $c > \left(1 + \frac{\sqrt{2}}{2}\right)\sqrt{2}$,用 Zoutendijk 的可行方向法产生迭代点列 $\{x^{(k)}\}$,则当 $k \to \infty$ 时,

$$x^{(k)} \to \hat{x} = \left(0,0,c - \sqrt{a}\left(1 + \frac{\sqrt{2}}{2}\right)\right)^T \neq x^*.$$

为此,1966 年,Zoutendijk 又提出"反锯齿措施"的可行方向法,称为"ε-紧约束可行方向法",该算法对非线性约束问题出现的"锯齿措施"得到了有效的遏制.

第三,对投影梯度法的改进

通过投影算子产生可行方向或近似可行方向，是把无约束优化方法推广到约束优化问题的另一类重要方法.英国数学家 Rosen 的投影梯度法是解决线性约束问题的首创，但该方法存在两个问题.

一个问题是收敛性没有得到证明.1969 年 E. Polak 对 Rosen 方法提出了相当复杂的修正，并证明当目标函数为凸函数和某些正则条件下，该修正算法收敛到最优解；1978 年中国科学院应用数学所桂湘云、章祥荪、赖炎连等人对 Polak 方法作了改进，使之在较弱的条件下收敛；1979 年，桂湘云、吴方、赖炎连对 Polak 方法作了进一步改进，并在目标函数 $f(x)$ 为凸的条件下证明了改进算法的收敛性.

另一个问题则是对非线性约束问题，Rosen 方法所产生的投影梯度一般说来不再是可行方向.1979 年我国数学家张承熙与夏人伟对 Rosen 方法提出一个改进算法，该算法将梯度方向与投影梯度方向进行线性组合，使之成为可行方向，即所谓的饶边现象(Hemspitching)，可导致算法的缓慢收敛.

无约束最优化方法习题

第1、2章习题

1. 画出下列函数的等值线，并指出函数值为正和负的区域；若有极小点，则指出它的坐标.

(1) $f(x_1,x_2)=3x_1^2+3x_2^2-6x_1+5$;

(2) $f(x_1,x_2)=x_1x_2$;

(3) $f(x_1,x_2)=x_1^2-x_2^2$;

(4) $f(x_1,x_2)=x_1^2+x_2^2$.

2. 把下列二次函数写成矩阵形式，并求各函数的极小点及其极小值.

(1) $f(x_1,x_2)=x_1^2+2x_2^2+x_1x_2$;

(2) $f(x_1,x_2,x_3)=3x_1^2+4x_2^2+x_3^2-2x_1x_2+3x_2x_3-4x_1+5x_3+10$.

3. 求下列函数的梯度及在指定点的梯度值.

(1) $f(x_1,x_2)=3x_1^2-4x_1x_2+x_2^2$，在 $\boldsymbol{x}=(0,1)^{\mathrm{T}}$.

(2) $f(x_1,x_2,x_3)=x_1^2+x_2^2+x_3^2-2x_1x_2+2x_2x_3+3x_2$，在 $\boldsymbol{x}=(1,-1,3)^{\mathrm{T}}$.

4. 求下列函数的梯度和 Hessian 矩阵：

(1) $f(\boldsymbol{x})=x_1^2+2x_2^2+3x_3^2-4x_1x_3$;

(2) $f(\boldsymbol{x})=3x_1x_2^2+e^{x_1x_2}$;

(3) $f(\boldsymbol{x})=\ln(x_1^2+x_1x_2+x_2^2)$;

(4) $f(\boldsymbol{x})=x_1x_2+\ln(x_1x_2)$.

5. 设 $\boldsymbol{Q}$ 是 $n\times n$ 对称阵，$\boldsymbol{x}$ 是 n 维向量，试证：

(1) $\nabla(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})=2\boldsymbol{x}$;

(2) $\nabla(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x})=2\boldsymbol{Q}\boldsymbol{x}$;

(3) $\nabla\boldsymbol{x}=\boldsymbol{I}$;

(4) $\nabla(\boldsymbol{Q}\boldsymbol{x})=\boldsymbol{Q}$.

6. 把函数 $f(x,y)=\sin x\sin y$ 在点 $\left(\frac{\pi}{4},\frac{\pi}{4}\right)^{\mathrm{T}}$ 处的邻域内展成 Taylor 级数，直到二次项为止.

7. 试证 $\boldsymbol{x}^*=(1,1)^{\mathrm{T}}$ 是函数

$$f(x_1,x_2)=(x_2-x_1^2)^2+(1+x_1)^2$$

的严格全局极小点，并应用 Laylor 展开式求这个函数在 $\boldsymbol{x}^*$ 处的邻域内的二次函数近似式.

8. 试证 $\boldsymbol{x}^*=(0,0)^{\mathrm{T}}$ 是函数

$$f(x_1,x_2)=4x_1^2+x_2^2-x_1^2x_2$$

的严格全局极小点，而 $x=(-2\sqrt{2},4)^T$ 和 $y=(2\sqrt{2},4)^T$ 是鞍点.

9. 设 f: $\mathbf{R}^n\to\mathbf{R}^1$ 定义如下：

$$f(x)=\frac{1}{2}x^TAx-b^Tx+c,$$

其中 A 为 $n\times n$ 对称阵，若 x^* 是线性方程组 $Ax=b$ 的解，试证：

(1) A 为正定时，x^* 是 $f(x)$ 的一个严格局部极小点；

(2) A 为负定时，x^* 是 $f(x)$ 的一个严格局部极大点.

10. 设 f_i: $\mathbf{R}^n\to\mathbf{R}^1(i=1,\cdots,m)$ 均为凸集 $D\subset\mathbf{R}^n$ 上的凸函数，证明 $f(x)=\sum_{i=1}^m f_i(x)$ 也是凸集 D 上的凸函数.

11. 设 f: $\mathbf{R}^n\to\mathbf{R}^1$ 是凸集 $D\subset\mathbf{R}^n$ 上的凸函数，且令 $S=\{x\in D\mid f(x)\geqslant a\}$，其中 a 为实数，试证明 S 为凸集.

12. 设 f: $\mathbf{R}^n\to\mathbf{R}^1$ 定义为

$$f(x)=(x_1-1)^2x_2+(x_1-1)^2x_1,$$

D 是含点 $x^*=(1,1)^T$ 的凸集，试证

(1) $f(x)$ 在 D 内是严格凸的；

(2) x^* 是 $f(x)$ 在 D 内的严格全局极小点.

13. 设 f: $\mathbf{R}^n\to\mathbf{R}^1$ 为 9 题中所定义的二次函数，试证明

(1) 若 $f(x)$ 是凸函数，则 A 半正定；

(2) 若 $f(x)$ 是严格凸函数，则 A 为正定.

14. 设目标函数 $f(x_1,x_2)=x_1^2+5x_2^2$，现在从点 $x_0=(1,1)^T$ 出发沿 $p=(-2,-1)^T$ 的方向作直线搜索，试求搜索到的极小点 z，并验证 $\nabla f(z)^Tp=0$.

15. 设 n 元二次函数

$$f(x)=\frac{1}{2}x^TQx+b^Tx+c,$$

其中 Q 为 $n\times n$ 对称正定矩阵，试证：从任意点 x_0（但 $\nabla f(x_0)\neq 0$）出发沿 $p=-Q^{-1}\nabla f(x_0)$ 的方向对 $f(x)$ 作直线搜索所得到的极小点 z 恰是 $f(x)$ 的极小点，而且最优步长因子等于 1.

16. 设 $\mathbf{R}^n$ 中的序列 $\{x_k\}$ 收敛于 x^*，若存在一个 K 和 k 无关的 $\beta\in(0,1)$ 使得对于 $k>K$ 有

$$\|x_{k+1}-x^*\|\leqslant\beta\|x_k-x^*\|,$$

或

$$\|x_{k+1}-x^*\|\leqslant\beta^k\|x_0-x^*\|,$$

试证 $\{x_k\}$ 是线性收敛的.

17. 试证

(1) $x_k=aq^k(|q|<1,a\neq 0)$ 为线性收敛；

(2) $x_k=\left(\frac{1}{k}\right)^k$ 为超线性收敛；

(3) $x_k = q^{2^k}$ ($|q| < 1$) 为二阶收敛.

18. 设 $\mathbf{R}^n$ 中的序列 $\{x_k\}$ 收敛于 x^*，若存在 $K \in (0,\infty)$ 及 $p>1$（与 k 无关）使得

$$\|x_{k+1} - x^*\| \leq K\|x_k - x^*\|^p \quad (\forall k \geq 0),$$

试证 $\{x_k\}$ 超线性收敛于 x^*.

第 3 章习题

1. 求一元函数 $\varphi(t) = (t-3)\sqrt{t}$ 的极小点，要求如下：

(1) 从 $t_0 = 0$ 出发，以 $h = 0.1$ 为步长确定一个搜索区间；

(2) 用二分法求极小点，误差限取为 $\varepsilon = 0.1$

2. 求一元函数 $\varphi(t) = -te^{-t}$ 的极小点，要求如下：

(1) 从 $t_0 = 0$ 出发，以 $h = 1$ 为步长确定一个搜索区间；

(2) 用 Newton 切线法求极小点，误差限取为 $\varepsilon = 0.1$

3. 求一元函数 $\varphi(t) = t^{2/3} - (t^2+1)^{1/3}$ 的极小点，要求如下：

(1) 从 $t_0 = 0$ 出发，以 $h = 0.1$ 为步长确定一个搜索区间；

(2) 用 0.618 法求极小点，误差限取为 $\varepsilon = 0.1$

4. 设 $f: \mathbf{R}^n \to \mathbf{R}^1$ 二次可微，g 是它的梯度，正定矩阵 G 是它的 Hessian 矩阵

(1) 试证把最速下降法用于 f 上时，第 k 次迭代的最优步长因子 α_k 可用下式近似：

$$\alpha_k = g_k^{\mathrm{T}} g_k / g_k^{\mathrm{T}} G_k g_k,$$

由此引出不用一维寻查的迭代公式

$$x_{k+1} = x_k - \frac{g_k^{\mathrm{T}} g_k}{g_k^{\mathrm{T}} G_k g_k} g_k \quad (k = 0,1,\cdots);$$

(2) 证明上述给出的算法是下降算法；

(3) 用上述公式解 $\min\{(x_1-1)^4 + 2x_2^2\}$，取初始点 $x_0 = (0,1)^{\mathrm{T}}$，迭代三次

5. 设 $\varphi(t) = f(x + tp)$， $f: \mathbf{R}^n \to \mathbf{R}^1$， $\varphi: \mathbf{R}^1 \to \mathbf{R}^1$，

证明

$$\varphi'(t) = \nabla f(x + \alpha p)^{\mathrm{T}} p,$$

$$\varphi''(t) = p^{\mathrm{T}} \nabla^2 f(x + \alpha p) p.$$

若

$$f(x) = x_2^2 - x_1^3 - 3x_1^2,$$

试确定计算 α_k 的近似公式，使 α_k 满足

$$f(x_k + \alpha_k p_k) = \min f(x_k + \alpha p_k).$$

第 4、5、6 章习题

1. 用最速下降法求解

$$\min\{x_1^2 + 2x_2^2\}.$$

设初始点为 $x_0 = (4,4)^{\mathrm{T}}$，迭代三次，并验证相邻两次迭代的寻查方向是相互垂直的（注：对于该问题，最优步长因子可以用公式表达）.

2. 试用 Newton 法求下列函数的极小点：

(1) $x_1^2+4x_2^2+9x_3^2-2x_1+18x_3$;

(2) $x_1^2-2x_1x_2+\frac{3}{2}x_2^2+x_1-2x_2$;

(3) $(x_1-1)^4+2x_2^2$.

以上第(1)、(2)小题的初始点可以任意选取，而第(3)小题的初始点取为 $\boldsymbol{x}_0=(0,1)^{\mathrm{T}}$.

3. 试用改进 Newton 法求解

$$\min\{4x_1^2+x_2^2-x_1^2x_2\},$$

取初始点 $\boldsymbol{x}_0=(0,4)^{\mathrm{T}}$，迭代三次.

4. 设 $\boldsymbol{A}=\begin{pmatrix}2&1\\1&2\end{pmatrix}$.

(1) 试证 $\boldsymbol{p}_0=(1,0)^{\mathrm{T}}, \boldsymbol{p}_1=(1,-2)^{\mathrm{T}}$ 线性无关且 A－共轭;

(2) 将 $\boldsymbol{v}=(1,1)^{\mathrm{T}}$ 表示为 $\boldsymbol{p}_0, \boldsymbol{p}_1$ 的线性组合;

(3) 验证 $\boldsymbol{p}_0=(1,0)^{\mathrm{T}}, \boldsymbol{p}_1=(1,1)^{\mathrm{T}}$ 线性无关但非 A－共轭.

5. 设 $\boldsymbol{A}$ 是 $n\times n$ 实对称阵，有互异的特征值，试证明 $\boldsymbol{A}$ 的特征向量为 $\boldsymbol{A}$－共轭.

6. 设 f: $\mathbf{R}^2\to\mathbf{R}^1$ 定义为

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x},$$

其中 $\boldsymbol{A}=\begin{pmatrix}2&1\\1&2\end{pmatrix}, \boldsymbol{b}=(3,3)^{\mathrm{T}}$.

试验证从 $\boldsymbol{x}_0=(0,0)^{\mathrm{T}}$ 出发，利用算法 5.1 沿 $\boldsymbol{p}_0=(1,0)^{\mathrm{T}}, \boldsymbol{p}_1=(1,-2)^{\mathrm{T}}$ 所生成的 $\boldsymbol{x}_2$ 是 $f(\boldsymbol{x})$ 的极小点.

7. 设 $\boldsymbol{A}$ 是 $n\times n$ 实对称阵，$\boldsymbol{v}_k(k=0,1,\cdots,n-1)$ 是线性无关的，又设 $\boldsymbol{p}_k(k=0,1,\cdots,n-1)$ 由以下递推公式产生：

$$\boldsymbol{p}_0=\boldsymbol{v}_0,\quad \boldsymbol{p}_{k+1}=\boldsymbol{v}_{k+1}-\sum_{i=0}^{k}\frac{\boldsymbol{v}_{k+1}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_i}{\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_i}\boldsymbol{p}_i,$$

证明 $\boldsymbol{p}_k(k=0,1,\cdots,n-1)$ 是 $\boldsymbol{A}$－共轭的.

8. 对于二次函数 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$，其中

$$\boldsymbol{A}=\begin{pmatrix}2&1&-1\\1&4&0\\-1&0&2\end{pmatrix},\ \boldsymbol{b}=(-1,1,2)^{\mathrm{T}},$$

试生成三个互为 $\boldsymbol{A}$－共轭的向量 $\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2$ 然后用算法 5.1 求 $f(\boldsymbol{x})$ 的极小点.

9. 设 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{v}\in\mathbf{R}^n$

设

$$\begin{cases}\boldsymbol{u}_1=\boldsymbol{x}_1+\alpha_1\boldsymbol{v},\\ \alpha_1:\ f(\boldsymbol{x}_1+\alpha_1\boldsymbol{v})=\min f(\boldsymbol{x}_1+\alpha\boldsymbol{v});\end{cases}$$

$$\begin{cases}\boldsymbol{u}_2=\boldsymbol{x}_2+\alpha_2\boldsymbol{v},\\ \alpha_2:\ f(\boldsymbol{x}_2+\alpha_2\boldsymbol{v})=\min f(\boldsymbol{x}_2+\alpha\boldsymbol{v});\end{cases}$$

$$w = u_2 - u_1.$$

试证 w 与 v 是 A - 共轭的.

10. 用算法 5.2 或算法 5.3 解

$$\min(x_1^2 - x_1x_2 + x_2^2 + 2x_1 - 4x_2),$$

取初始点 $x_0 = (2,2)^T$,迭代三次.

11. 用算法 5.4(PRP 算法)解

$$\min\{(1 - x_1)^2 + 2(x_2 - x_1^2)^2\},$$

取初始点 $x_0 = (0,0)^T$,迭代三次.

12. 有一种改进的最速下降法,称为平行线法或 Partan 方法,其算法如下:

已知:目标函数 $f(x)$ 及其梯度 $\nabla f(x)$.

① 选定初始点 x_0,作直线搜索

$$\begin{cases} x_1 = x_0 + \alpha_0 p_0, \\ p_0 = -\nabla f(x_0), \\ \alpha_0 : f(x_0 + \alpha_0 p_0) = \min f(x_0 + \alpha p_0), \end{cases}$$

计算 $f_1 = f(x_1)$;置 $k = 1$.

② 作直线搜索

$$\begin{cases} Z_k = x_k + \alpha_k p_k, \\ p_k = -\nabla f(x_k), \\ \alpha_k : f(x_k + \alpha_k p_k) = \min f(x_k + \alpha p_k). \end{cases}$$

$$\begin{cases} x_{k+1} = Z_k + \beta_k p_k, \\ p_k = Z_k - x_{k-1}, \\ \beta_k : f(Z_k + \beta_k p_k) = \min f(Z_k + \beta p_k), \end{cases}$$

计算 $f_{k+1} = f(x_{k+1})$

③ 若 H - 终止准则满足,则 x_{k+1} 和 f_{k+1} 就是最优解,打印,停机;否则,置 $k = k + 1$ 转②

(1) 试对平行切线法作几何解释;

(2) 试用平行切线法求解本习题的第 1 题,并说明为什么经过两次迭代就求到了极小点.

13. 有一种共轭方向法,称为 Powell 方法,其算法如下:

已知:正定二次函数 $f(x)$

① 给定初始点 x_0,令 $k = 0$;

② 令 $p_i = e_i (i = 1, \cdots, n)$;

③ 对 $i = 1, \cdots, n$ 计算

$$\begin{cases} x_i = x_{i-1} + \alpha_i p_i, \\ \alpha_i : f(x_{i-1} + \alpha_i p_i) = \min f(x_{i-1} + \alpha p_i); \end{cases}$$

④ 令 $p = x_n - x_0$

$$\begin{cases} x_{n+1} = x_n + \alpha_{n+1} p, \\ \alpha_{n+1} : f(x_n + \alpha_{n+1} p) = \min f(x_n + \alpha p_i); \end{cases}$$

⑤ 检查 $k = n-1$?

是令 $x^* = x_{n+1}$，停止计算.

⑥ 令 $p_i = p_{i+1}(i = 1,2,\cdots,n-1)$,

$$p_n = p;$$

⑦ 令 $x_0 = x_{n+1}$，$k = k+1$ 转③.

试就 $f(x) = f(x_1, x_2, x_3)$ 的情形，证明当 $k=2$ 时，由上述算法构造的寻查方向 p_2, p_3，p 为一组共轭方向，x_4 为 $f(x)$ 的极小点.

14. 分别利用 SR1 方法和 DFP 方法求

$$\min(x_1^2 - x_1x_2 + x_2^2 + 2x_1 - 4x_2).$$

初始点取为 $x_0 = (2,2)^T$，初始矩阵取为单位矩阵，并验证每个算法所生成的两个搜索方向都是关于矩阵 $Q = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ 共轭的.

15. 用 DFP 方法求解

$$\min\{(1-x_1)^2 + 2(x_2 - x_1^2)^2\},$$

初始点取为 $x_0 = (0,0)^T$，初始矩阵取为单位矩阵，迭代二次.

第 7、8、9 章习题

1. 用 Powell 方法求解下列问题

(1) $\min\{x_1^2 + 4x_2^2\}$，取 $x_0 = (2,2)^T$;

(2) $\min\{17x_1^2 + 12x_1x_2 + 8x_2^2\}$ $x_0 = (2,2)^T$.

2. 求超定方程组 $Ax = y$ 的最小二乘解

其中
$$A = \begin{pmatrix} 2 & -1 \\ 1 & 1 \\ 3 & 2 \end{pmatrix}, \quad y = (4,3,6)^T.$$

3. 用最小二乘原理求一个形如 $y = a + bx^2$ 的经验公式，使其与下列数据相拟合

x_i	19	25	31	38	44
y_i	19.0	32.3	49.0	73.3	97.8

4. 用最小二乘原理求一个形如 $y = ae^{bx}$ 的经验公式，使其与下列数据相拟合

x_i	1	2	3	4
y_i	60	30	20	10

5. 设一物体作匀加速直线运动，观测数据如下：

t(s)	0	0.9	1.9	3.0	3.9	5.0
s(m)	0	10	30	50	80	110

试求该物体的初速度及加速度.

6. 设超定方程组

$$\begin{pmatrix} A_{m\times n} \\ \text{—} \\ \lambda D \end{pmatrix} x = \begin{pmatrix} f_{m\times 1} \\ \text{—} \\ O_{n\times 1} \end{pmatrix}$$

其中 λ 为常数，$\boldsymbol{D} = \mathrm{diag}(d_{11}, d_{22}, \cdots, d_{nn})$，

试证明其最小二乘解 x 满足

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \lambda\boldsymbol{D})\boldsymbol{x} = -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{f}.$$

7. 用 Golub 方法求 $\boldsymbol{Ax} = \boldsymbol{y}$ 的最小二乘解，

其中
$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{pmatrix}, \ \boldsymbol{y} = (6, 0.01, 0.02, 0.03)^{\mathrm{T}}.$$

8. 设 $\boldsymbol{A} = (a_{ij})_{m\times n}$，rank $(\boldsymbol{A}) = n$，

试证明 $A^{\mathrm{T}}\boldsymbol{A} = R^{\mathrm{T}}R$（$R$ 为 n 阶上三角阵），并设计计算$(A^{\mathrm{T}}A)^{-1}$的计算方法.

9. 设有非线性方程组

$$f_1(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1 = 0,$$
$$f_2(\boldsymbol{x}) = 2x_1 + x_2 - 2 = 0.$$

（1）列出求解这个方程组的非线性最小二乘问题的数学模型；

（2）写出求解该问题的 Gauss-Newton 法迭代公式的具体形式；

（3）初始点取为 $\boldsymbol{x}_0 = (2,2)^{\mathrm{T}}$，迭代三次.

10. 已知某物理量 y 与另两个物理量 t_1, t_2 的关系为

$$y = \frac{x_1 x_3 t_1}{1 + x_1 t_1 + x_2 t_2},$$

其中 x_1, x_2, x_3 是待定参数，为确定这三个参数测得 t_1, t_2 和 y 的分组数据

t_1	1.0	2.0	1.0	2.0	0.1
t_2	1.0	1.0	2.0	2.0	0.0
y	0.126	0.219	0.076	0.126	0.186

（1）用最小二乘法建立关于确定 x_1, x_2, x_3 的数学模型；

（2）对列出的非线性最小二乘问题写出 Gauss-Newton 法迭代公式的具体形式.

11. 设 $S: \mathbf{R}^2 \to \mathbf{R}^1$ 定义如下

$$S(\boldsymbol{x}) = f(\boldsymbol{x})^{\mathrm{T}} f(\boldsymbol{x}),$$

其中 $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}))^{\mathrm{T}}$, $f_1(\boldsymbol{x}) = x_2 - x_1^2$, $f_2(\boldsymbol{x}) = 1 - x_1$.

对于任何的初始点 x_0，试写出按如下 Newton 迭代公式求极小点的计算步骤.

$$\begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k, \\ \boldsymbol{p}_k = -[\nabla^2 s(\boldsymbol{x}_k)]^{-1} s(\boldsymbol{x}_k), \\ \alpha_k \cdot s(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) = \min s(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k), \end{cases} \quad k = 0, 1, \cdots,$$

并证明　当 $\boldsymbol{x} \to \boldsymbol{x}^*$ 时，$\nabla^2 s(\boldsymbol{x}) \to 2\boldsymbol{A}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{A}(\boldsymbol{x})$，

其中
$$\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{D}f(\boldsymbol{x}).$$

下　　篇
约束最优化方法

第 10 章　线性规划及其解法

目前,线性规划在相关的应用领域中是一种被广泛使用又卓有成效的约束优化方法,同时它的求解方法也为某些非线性规划求解算法的设计提供了帮助.早在 20 世纪 30 年代,Kantolovch 等在研究生产管理和运输等方面就开始应用线性规划的优化方法.40 年代末 Dantzig 等人提出单纯形法,从面为线性规划奠定了理论基础.本章主要讨论线性规划的基本概念、基本理论和求解方法.

10.1　线性规划问题举例

例 10.1　生产计划问题

设某企业现有 m 种资源 $A_i(i=1,2,\cdots,m)$ 用于生产 n 种产品 $B_j(j=1,2,\cdots,n)$,每种资源的拥有量和每种产品所消耗的资源量,以及单位产品的利润如表 10-1,试问如何安排生产计划使得该企业获利最大?

表 10-1

资源＼产品	B_1	B_2	$\cdots$	B_n	资源总量
A_1	a_{11}	a_{12}	$\cdots$	a_{1n}	b_1
A_2	a_{21}	a_{22}	$\cdots$	a_{2n}	b_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_m	a_{m1}	a_{m2}	$\cdots$	a_{mn}	b_m
单位利润	c_1	c_2	$\cdots$	c_n	

解　设产品 B_j 产量为 $x_j(j=1,2,\cdots,n)$，称之为决策变量，所得的利润为 z，则要解决的问题的目标是使得（利润）函数 $z=\sum_{j=1}^{n}c_jx_j$ 有最大值. 决策变量所受的约束条件为

$$\begin{cases}\sum_{j=1}^{n}a_{ij}x_j\leqslant b_i & (i=1,2,\cdots,m),\\ x_j\geqslant 0 & (j=1,2,\cdots,n).\end{cases}$$

于是问题可归结为求目标函数在约束条件下的最大值问题. 显然目标函数和约束条件都是决策变量的线性函数，即有下面的线性规划模型

$$\max z=\sum_{j=1}^{n}c_jx_j,$$

$$\text{s. t.}\begin{cases}\sum_{j=1}^{n}a_{ij}x_j\leqslant b_i & (i=1,2,\cdots,m),\\ x_j\geqslant 0 & (j=1,2,\cdots,n).\end{cases}$$

例 10.2　配料问题

每种食品含有特定种类和比例的营养成分，从医学上知道每人每天对每种营养成分的最低需求量. 现设每人每天所需的 m 种营养成分为 A_i $(i=1,2,\cdots,m)$，每种营养成分的最低需求量分别为 $b_i(i=1,2,\cdots,m)$，现市场上有 n 种不同的食品 $B_j(j=1,2,\cdots,n)$，每种食品的单价为 $c_j(j=1,2,\cdots,n)$，每单位 B_j 含 A_i 的量为 a_{ij}（如表 10－2 所示）. 问如何制定采购计划，既可以保证每人的营养需要，又使每天的费用最低？

表 10－2

营养成分 \ 食品	B_1	B_2	$\cdots$	B_n	营养成分最低需求量
A_1	a_{11}	a_{12}	$\cdots$	a_{1n}	b_1
A_2	a_{21}	a_{22}	$\cdots$	a_{2n}	b_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_m	a_{m1}	a_{m2}	$\cdots$	a_{mn}	b_m
食品单价	c_1	c_2	$\cdots$	c_n	

解　设决策变量 x_j 表示食品 B_j 的采购量，目标函数为所需的总费用 $z=\sum_{j=1}^{n}c_jx_j$，约束条件为

$$\sum_{j=1}^{n}a_{ij}x_j\geqslant b_i\quad(i=1,2,\cdots,m),$$

$$x_j \geqslant 0 \quad (j = 1,2,\cdots,n).$$

该问题的线性规划模型为

$$\min z = \sum_{j=1}^{n} c_j x_j,$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{n} a_{ij} x_j \geqslant b_i & (i = 1,2,\cdots,m), \\ x_j \geqslant 0 & (j = 1,2,\cdots,n). \end{cases}$$

10.2　线性规划问题的基本概念及解的性质

10.2.1　线性规划模型的一般形式

由实际问题建立起来的线性规划模型有多种不同形式，如目标函数可能是最大化问题，也可能是最小化问题；约束条件可能是"≤"，也可能是"≥"，还可能是"="；决策变量可能有非负要求，也可能没有符号要求. 于是可以得到线性规划模型的一般形式为

$$\max\ (\min) z = \sum_{j=1}^{n} c_j x_j,$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{n} a_{ij} x_j \leqslant (\geqslant, =) b_i & (i = 1,2,\cdots,m), \\ x_j \geqslant 0 & (j = 1,2,\cdots,n), \end{cases} \tag{10-1}$$

也可表示为矩阵形式

$$\max\ (\min) z = \boldsymbol{c}^{\mathrm{T}} \boldsymbol{x},$$

$$\text{s.t.} \begin{cases} \boldsymbol{A}\boldsymbol{x} \leqslant (\geqslant, =) \boldsymbol{b}, \\ \boldsymbol{x} \geqslant 0. \end{cases}$$

向量形式

$$\max\ (\min) z = \boldsymbol{c}^{\mathrm{T}} \boldsymbol{x},$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{n} \boldsymbol{p}_j x_j \leqslant (\geqslant, =) \boldsymbol{b}, \\ \boldsymbol{x} \geqslant 0. \end{cases} \tag{10-2}$$

其中 $\boldsymbol{c}^{\mathrm{T}} = (c_1, c_2, \cdots, c_n)$ 称其为目标函数的系数向量；$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 称其为决策向量；$\boldsymbol{b} = (b_1, b_2, \cdots, b_m)^{\mathrm{T}}$ 称其为约束方程组的常数向量；$\boldsymbol{A} = (a_{ij})_{m \times n}$ 称其为约束方程组的系数矩阵；$\boldsymbol{p}_j = (a_{1j}, a_{2j}, \cdots, a_{mj})^{\mathrm{T}} (j = 1, 2, \cdots, n)$ 称其为约束方程组的系数向量.

为了方便问题的讨论，通常我们规定一种标准形式，实际中非标准形

式的线性规划模型都可以化为标准型来求解. 线性规划模型的标准型规定为

$$\max z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

$$\text{s.t.}\begin{cases}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\\ \boldsymbol{x} \geqslant 0.\end{cases} \tag{10-3}$$

对于非标准型的线性规划模型都可以化为标准型,其方法如下:

(1) 目标函数为最小化问题:令 $z' = -z$,则 $\max z' = -\min z = -\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$;

(2) 约束条件为不等式:对于不等号"≤(≥)"的约束条件,则可在"≤(≥)"的左端加上(或减去)一个非负变量(称为松弛变量)使其变为等式.

(3) 对于无约束的决策变量:譬如 $x \in (-\infty, +\infty)$,则令 $x = x' - x''$,使得 $x', x'' \geqslant 0$,代入模型即可.

例 10.3　将下述线性规划问题化为标准型

$$\min z = -x_1 + 2x_2 - 3x_3,$$

$$\text{s.t.}\begin{cases}x_1 - x_2 + x_3 \geqslant 2,\\ -x_1 - x_2 - x_3 \geqslant -7,\\ -3x_1 + x_2 + 2x_3 = 5,\\ x_2, x_3 \geqslant 0, x_1 \text{ 无约束}.\end{cases}$$

解　(1) 令 $z' = -z$,把求 min 改为求 max;

(2) 用 $x_1' - x_1''$ 替换 x_1,其中 $x_1', x_1'' \geqslant 0$;

(3) 在第一个约束不等式≥号的左端减去一个剩余变量 x_4;

(4) 在第二个约束不等式中,首先两边同乘以 -1,得 $x_1 + x_2 + x_3 \leqslant 7$,然后在此不等式的左端加上一个松弛变量 x_5,则其标准型为

$$\max z' = (x_1' - x_1'') - 2x_2 + 3x_3 + 0x_4 + 0x_5,$$

$$\text{s.t.}\begin{cases}(x_1' - x_1'') - x_2 + x_3 - x_4 = 2,\\ (x_1' - x_1'') + x_2 + x_3 + x_5 = 7,\\ -3(x_1' - x_1'') + x_2 + 2x_3 = 5,\\ x_1', x_1'', x_2, x_3, x_4, x_5 \geqslant 0.\end{cases}$$

10.2.2　线性规划问题解的概念

对于线性规划问题

$$\max z = \sum_{j=1}^{n} c_j x_j, \tag{10-4}$$

$$\text{s. t.} \begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i \quad (i = 1,2,\cdots,m), & (10-5) \\ x_j \geq 0 \quad (j = 1,2,\cdots,n). & (10-6) \end{cases}$$

可行解：满足(全部)约束条件(10－5)式和(10－6)式的决策向量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 称为可行解.

可行域：全部可行解所构成的集合称为可行域，它是一个 n 维欧氏空间 $\mathbf{R}^n$ 中的点集，而且是一个"凸多面体"，也称为可行集.

最优解：使目标函数达到最优值的可行解，即满足(10－4)式的可行解称为最优解.

无界解：对于最大化问题，若使得目标函数在可行域中无上界；对于最小化问题，若使得目标函数在可行域中无下界，则称该问题为无界解.

基矩阵：如果系数矩阵 $\boldsymbol{A}$ 是 $m \times n$ 矩阵，且秩为 $R(\boldsymbol{A}) = m$，则称任意一个 $m \times m$ 阶非奇异子矩阵 $\boldsymbol{B}$ 为线性规划问题的基矩阵，简称为基. 不失一般性，可设

$$\boldsymbol{B} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_m),$$

称 $\boldsymbol{p}_i (i = 1,2,\cdots,m)$ 为基向量，与基向量 $\boldsymbol{p}_i$ 对应的变量 $x_i (i = 1,2,\cdots, m)$ 为基变量.

基解：如果问题的基为 $\boldsymbol{B}$，对应的基变量为 $x_i (i = 1,2,\cdots,m)$，令非基变量 $x_{m+1} = x_{m+2} = \cdots = x_n = 0$，此时所形成的方程组(10－5)为基方程组. 因为 $\det \boldsymbol{B} \neq 0$，由克拉默法则可解出唯一的解

$$\boldsymbol{x}_B = (x_1, x_2, \cdots, x_m)^{\mathrm{T}},$$

则称

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_m, 0, 0, \cdots, 0)^{\mathrm{T}}$$

为线性规划问题的基解.

基可行解：满足非负条件(10－6)的基解称为基可行解.

退化基解：基解中有一个或多于一个基变量为 0 时，这个解称为退化基解，否则称为非退化基解.

退化基可行解：若基可行解中有一个或多于一个基变量为零时，这个解称为退化基可行解.

10.2.3　线性规划问题解的性质

定理 10.1　若线性规划问题(10－3)存在可行域，则其可行域是凸集.

证　设线性规划问题(10－3)，其可行域为 $R=\{x\mid Ax=b,x\geqslant 0\}$，从 R 中任取两点 $x^{(1)}$ 和 $x^{(2)}(x^{(1)}\neq x^{(2)})$，则

$$Ax^{(1)}=b, Ax^{(2)}=b, x^{(1)}\geqslant 0, x^{(2)}\geqslant 0.$$

令 x 为连接 $x^{(1)}$ 和 $x^{(2)}$ 的线段上的任一点，即有

$$x=\mu x^{(1)}+(1-\mu)x^{(2)}\ (0\leqslant\mu\leqslant 1),$$

则

$$\begin{aligned}Ax&=A[\mu x^{(1)}+(1-\mu)x^{(2)}]=\mu Ax^{(1)}+Ax^{(2)}-\mu Ax^{(2)}\\&=\mu b+b-\mu b=b,\end{aligned}$$

又因 $x^{(1)}\geqslant 0, x^{(2)}\geqslant 0$ 及 $0\leqslant\mu\leqslant 1$，故 $x\geqslant 0$，即 $x\in R$.

根据第 1 章凸集的定义(定义 1.1)知，可行域 R 为凸集.

引理 10.1　线性规划问题的可行解 $x=(x_1,x_2,\cdots,x_n)^{\mathrm T}$ 为基可行解的充要条件是 x 的正分量所对应的系数列向量组是线性无关的.

证　必要性：因 x 是基本可行解，故 x 的正分量就是各个基变量，而各个基变量对应的系数列向量就是各个基向量. 由基的定义，它们是线性无关的.

充分性：若向量 $p_1,p_2,\cdots,p_k$ 线性无关，则必有 $k\leqslant m$. 当 $k=m$ 时，它们恰构成一个基，从而可行解 $x=(x_1,x_2,\cdots,x_k,0,0,\cdots,0)^{\mathrm T}$ 为相应的基可行解. 当 $k<m$ 时，则一定可以从其余的列向量中取出 $m-k$ 个与 $p_1,p_2,\cdots,p_k$ 构成最大的线性无关向量组，其对应的解恰为 x，由定义它是基可行解.

定理 10.2　线性规划问题的任一个基可行解 x 对应于可行域的一个顶点.

证　下面分两步来证明. 不失一般性，假设基可行解 x 的前 m 个分量为正，正分量所对应的系数列向量为 $p_1,p_2,\cdots,p_m$，则

$$\sum_{i=1}^{m}p_ix_i=b. \tag{10-7}$$

(1) 若 x 是基可行解，则一定是可行域的顶点.

反证，如果 $x=(x_1,x_2,\cdots,x_m,0,0,\cdots,0)^{\mathrm T}$ 是一个基可行解，但不是可行域的顶点.

事实上，由定理 10.1 知 x 必可以表示为可行域中两个不同的点

$$x^{(1)}=(x_1^{(1)},x_2^{(1)},\cdots,x_n^{(1)})^{\mathrm T},$$

$$x^{(2)}=(x_1^{(2)},x_2^{(2)},\cdots,x_n^{(2)})^{\mathrm T}$$

的凸组合. 即

$$x = \mu x^{(1)} + (1-\mu)x^{(2)} (0 < \mu < 1).$$

因为 x 的前 m 个分量为正,x 的后 $n-m$ 个分量为 0,又 $\mu>0, 1-\mu>0, x^{(1)}>0, x^{(2)}>0$,故 $x^{(1)}, x^{(2)}$ 的后 $n-m$ 个分量也为 0,即

$$\sum_{i=1}^{m} p_i x_i^{(1)} = b, \tag{10-8}$$

$$\sum_{i=1}^{m} p_i x_i^{(2)} = b. \tag{10-9}$$

(10－8)和(10－9)两式相减可得

$$\sum_{i=1}^{m} p_i (x_i^{(1)} - x_i^{(2)}) = b.$$

因 $x^{(1)} \neq x^{(2)}$,故 $x_i^{(1)} - x_i^{(2)}$ 不全为零. 由定义知 x 正分量对应的系数列向量 $p_1, p_2, \cdots, p_m$ 线性相关,故 x 不是基可行解. 这与假设矛盾,故命题(1)成立.

(2) 若 x 是可行域的顶点,则 x 一定是基可行解.

反证,如果 x 是可行域的顶点,但 x 不是基可行解.

事实上,由引理 10.1 知,若 x 不是基可行解,则向量组 $p_1, p_2, \cdots, p_m$ 必线性相关,故存在一组不全为零的数 $k_i (i=1,2,\cdots,m)$ 使

$$\sum_{i=1}^{m} p_i k_i = 0. \tag{10-10}$$

用一个不为 0 的数 μ 乘(10－10)式,再分别与(10－7)式相加和相减可得

$$\sum_{i=1}^{m} p_i (x_i + \mu k_i) = b,$$

$$\sum_{i=1}^{m} p_i (x_i - \mu k_i) = b.$$

令

$$x^{(1)} = (x_1 + \mu k_1, x_2 + \mu k_2, \cdots, x_k + \mu k_m, 0, 0, \cdots, 0)^T,$$

$$x^{(2)} = (x_1 - \mu k_1, x_2 - \mu k_2, \cdots, x_k - \mu k_m, 0, 0, \cdots, 0)^T.$$

则取 $x = \frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)}$ 为 $x^{(1)}$ 和 $x^{(2)}$ 的连线中点,而当 μ 充分小时,可以保证

$$x_i \pm \mu k_i \geqslant 0 \ (i = 1, 2, \cdots, m),$$

即 $x^{(1)}$ 和 $x^{(2)}$ 是可行解,故 x 不是可行域的顶点. 这与假设矛盾,故命题(2)成立.

引理 10.2　若 K 是有界凸集,则任意一点 $x \in K$ 均可表示为 K 的顶

点的凸组合.

证明略.

由引理 10.2 可知,线性规划问题的任一可行解均可表示为基可行解的凸组合.

定理 10.3 若线性规划问题(10-3)的可行域有界,则目标函数一定可以在其可行域的顶点上达到最优.

证 设 $x^{(1)},x^{(2)},\cdots,x^{(k)}$ 是可行域的顶点,而基可行解 $x^{(0)}$ 不是可行域的顶点,且目标函数在 $x^{(0)}$ 处达到最优值 $z^*=c^{\mathrm{T}}x^{(0)}$. 则由引理 10.2 知,$x^{(0)}$ 可表示为可行域顶点的凸组合,即

$$x^{(0)}=\sum_{i=1}^{k}\mu_i x^{(i)},$$

其中 $\mu_i\geqslant 0$ 且 $\sum_{i=1}^{k}\mu_i=1\ (i=1,2,\cdots,k)$,则

$$c^{\mathrm{T}}x^{(0)}=c^{\mathrm{T}}\sum_{i=1}^{k}\mu_i x^{(i)}=\sum_{i=1}^{k}\mu_i c^{\mathrm{T}}x^{(i)}.$$

假设存在某一顶点 $x^{(m)}$,其目标函数值 $c^{\mathrm{T}}x^{(m)}$ 是所有 $c^{\mathrm{T}}x^{(i)}\ (i=1,2,\cdots,k)$ 中的最大者. 将 $x^{(m)}$ 代替式中其余 $x^{(i)}$,得

$$\sum_{i=1}^{k}\mu_i c^{\mathrm{T}}x^{(i)}\leqslant\sum_{i=1}^{k}\mu_i c^{\mathrm{T}}x^{(m)}=c^{\mathrm{T}}x^{(m)},$$

即 $c^{\mathrm{T}}x^{(0)}\leqslant c^{\mathrm{T}}x^{(m)}$. 又因为 $c^{\mathrm{T}}x^{(0)}$ 是最大值,故只有 $c^{\mathrm{T}}x^{(0)}=c^{\mathrm{T}}x^{(m)}$,即目标函数在顶点 $x^{(m)}$ 处达到最大值.

一般说来,线性规划问题的可行域可能有三种情形:有界凸集、无界凸集和空集. 线性规划问题的最优解取决于问题可行域的结构,即可能有如下三种情形(以二维最大化问题为例).

1. 唯一有限最优解

此时无论可行域是有界还是无界,线性规划问题的最优解必在可行域的某个顶点处达到. 如图 10-1 所示.

2. 择一有限最优解

图 10-2 表明线性规划问题可能具有无穷多个有限最优解. 图 10-2a 表明,有界可行域的两个顶点 x_1^*,x_2^* 的连线上的所有点均为问题的有限最优解;图 10-2b 表明,在无界可行域的一个顶点 x^* 上,以它为起点的射线上的所有点均为问题的有限最优解.

3. 无界解

图 10-3 表明,可行域是无界的,且最优解也是无限的. 此时目标函数的等值线沿方向 $-c$ 无限地移动并且总是与可行域相交,此时最优值为 $+\infty$,称线性规划问题有无界解.

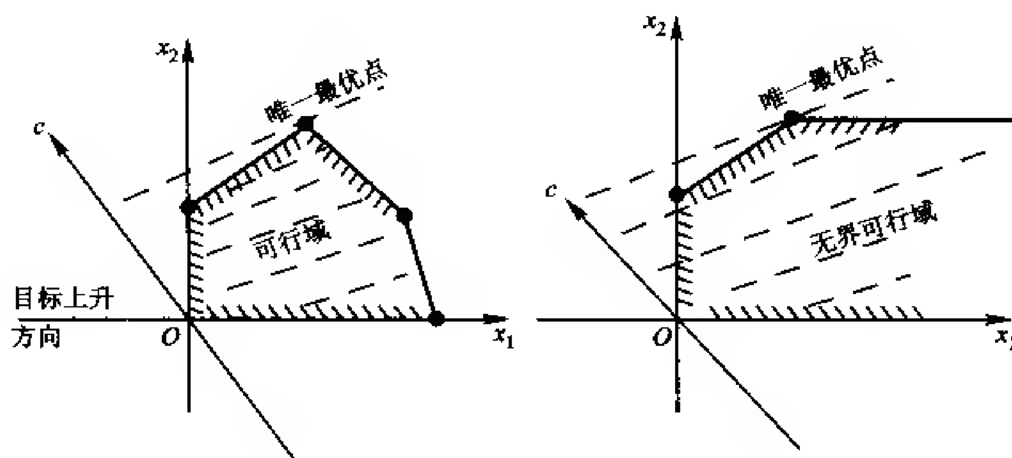


图 10－1

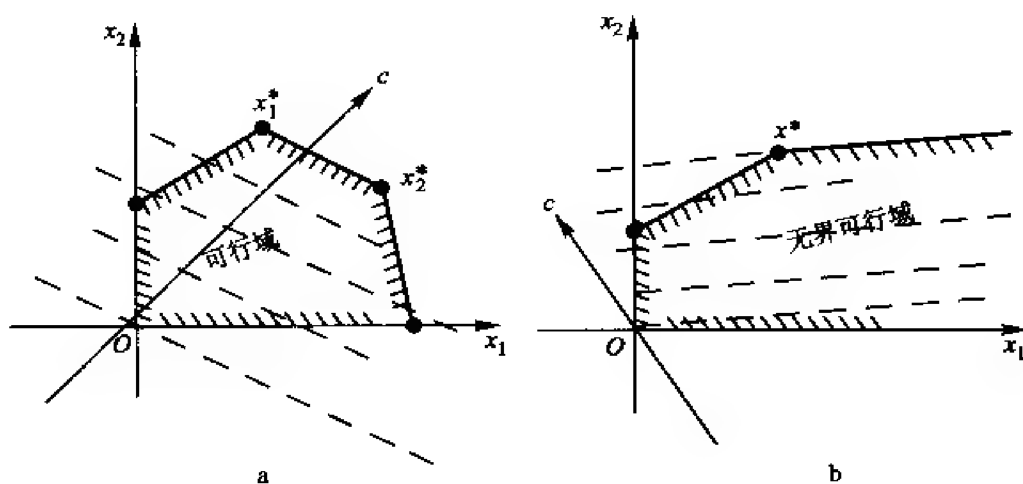


图 10－2

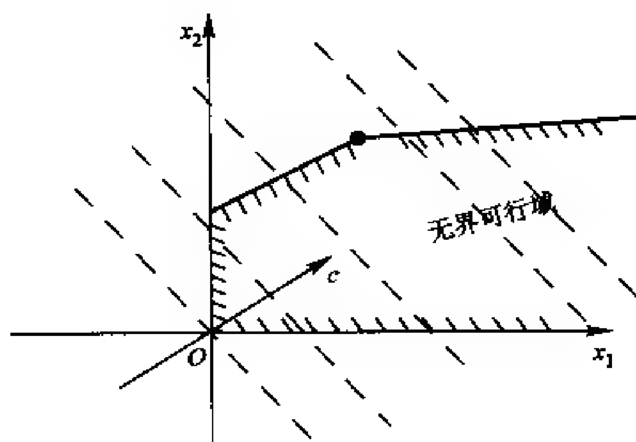


图 10－3

10.3　单纯形法

10.3.1　单纯形法原理

根据线性规划解的概念和性质,如果线性规划问题存在最优解,则一定有一个基可行解为最优解. 为此,求解线性规划可采用下面的方法:首先将线性规划问题化为标准型,在可行域中寻求一个基可行解(对应一个可行域的顶点);然后检查该基可行解是否为最优解,如果不是,则设法再求另一个没有检查过的基可行解,并且使目标函数不断增大. 如此进行下去,直到得到某一个基可行解为最优解为止. 由于可行域的顶点个数是有限的,则可以证明经过有限次迭代一定可以求得问题的最优解或判定出问题无最优解. 现在要解决的问题:如何求出第一个基可行解? 如何判断基可行解是否为最优解? 如何由一个基可行解过渡到另一个基可行解? 解决这些问题的方法称为单纯形法.

确定问题的初始基可行解是解决问题的第一步,即首先要找出初始可行基,即是由 m 个不同的单位列向量构成的一个 $m\times m$ 阶单位阵. 其方法如下:

如果线性规划问题的约束条件全部为等式,则在其系数矩阵中总可以找到一个 $m\times m$ 阶单位矩阵. 如果问题的所有约束条件均是"$\leqslant$"形式的不等式,则可用化标准型的方法在每个约束条件的左端加上一个松弛变量,从而可得到一个 $m\times m$ 阶单位矩阵. 如果约束条件为"$\geqslant$"形式的不等式及等式的情况,则采用人造基的方法,即引入人工变量以构成初始可行基,具体的在 10.3.2 节讨论. 综上所述,最后整理可得

$$\max z = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

$$= \sum_{i=1}^{m} c_ib_i - \sum_{j=m+1}^{n}\left(-c_j + \sum_{i=1}^{m} c_ia_{ij}\right)x_j,$$

$$\text{s.t.}\begin{cases} x_1 + \cdots + a_{1,m+1}x_{m+1} + \cdots + a_{1n}x_n = b_1, \\ x_2 + \cdots + a_{2,m+1}x_{m+1} + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots \\ x_m + a_{m,m+1}x_{m+1} + \cdots + a_{mn}x_n = b_m, \\ x_j \geqslant 0\ (j = 1,2,\cdots,n). \end{cases} \tag{10-11}$$

这样总能得到一个 $m\times m$ 单位矩阵

$$B=(p_1,p_2,\cdots,p_m)=\begin{pmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&1\end{pmatrix},$$

称之为线性规划问题的一个初始可行基,而对应的变量 $x_i(i=1,2,\cdots,m)$称为基变量,其余变量 $x_i(i=m+1,\cdots,n)$为非基变量.

令非基变量 $x_{m+1}=\cdots=x_n=0$,则得 $x_i=b_i(i=1,\cdots,m)$,可得初始基可行解为

$$x=(x_1,x_2,\cdots,x_m,0,\cdots,0)^{\mathrm T}=(b_1,b_2,\cdots,b_m,0,\cdots,0)^{\mathrm T}.$$

用单纯形法求解线性规划问题时,常用一种表上作业法,这种表格称为单纯形表.含初始基可行解的单纯形表称为初始单纯形表(如表 10-3).

表 10-3

$c_j\to$			c_1	c_2	$\cdots$	c_m	$\cdots$	c_j	$\cdots$	c_n	θ
c_B	x_B	b	x_1	x_2	$\cdots$	x_m	$\cdots$	x_j	$\cdots$	x_n	
c_1	x_1	b_1	1	0	$\cdots$	0	$\cdots$	a_{1j}	$\cdots$	a_{1n}	θ_1
c_2	x_2	b_2	0	1	$\cdots$	0	$\cdots$	a_{2j}	$\cdots$	a_{2n}	θ_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$	$\vdots$
c_m	x_m	b_m	0	0	$\cdots$	1	$\cdots$	a_{mj}	$\cdots$	a_{mn}	θ_m
σ_j			0	0	$\cdots$	0	$\cdots$	$c_j-\sum_{i=1}^{m}c_ia_{ij}$	$\cdots$	$c_n-\sum_{i=1}^{m}c_ia_{in}$	

对于线性规划问题(10-11),下面给出单纯形法的求解算法.

算法 10.1

(一) 初始步

给出初始单纯形表(表 10-3),对应的有

$$\begin{aligned}&x_1+\cdots+a_{1,m+1}^{(0)}x_{m+1}+\cdots+a_{1n}^{(0)}x_n=b_1^{(0)},\\&x_2+\cdots+a_{2,m+1}^{(0)}x_{m+1}+\cdots+a_{2n}^{(0)}x_n=b_2^{(0)},\\&\cdots\cdots\cdots\cdots\\&x_m+a_{m,m+1}^{(0)}x_{m+1}+\cdots+a_{mn}^{(0)}x_n=b_m^{(0)},\end{aligned}$$

由此得到一个初始的基可行解

$$x^{(0)} = (b_1^{(0)}, b_2^{(0)}, \cdots, b_m^{(0)}, 0, \cdots, 0)^T.$$

（二）迭代步

① 给出初始基可行解 $x^{(0)}$；

② 计算非基变量的判别数

$$\sigma_j = c_j - \sum_{i=1}^{m} c_i a_{ij}^{(0)} (j = m+1, \cdots, n);$$

③ 令 $\sigma_k = \max(\sigma_j)$；

若 $\sigma_k \leqslant 0, x^{(0)} \Rightarrow x^*$，即得到最优解，计算结束；

若 $\sigma_k > 0$，转④；

④ 检查 $a_{ik}^{(0)} \leqslant 0\ (i = 1, 2, \cdots, m)$ 是否成立？

是，问题无最优解，停止计算；

否，转⑤；

⑤ 计算 $\theta^{(0)} = \min\left\{\frac{b_i^{(0)}}{a_{ik}^{(0)}} \middle| a_{ik}^{(0)} > 0\right\} = \frac{b_l^{(0)}}{a_{lk}^{(0)}}$，

由此定出主列为第 k 列，主行为第 l 行，主元为 $a_{lk}^{(0)}$；

⑥ 执行 Gauss – Jordan 消元：

$$\begin{pmatrix} p_1 & \cdots & p_l & \cdots & p_m & p_{m+1} & \cdots & p_k & \cdots & p_n & b \\ 1 & \cdots & 0 & \cdots & 0 & a_{1,m+1}^{(0)} & \cdots & a_{1k}^{(0)} & \cdots & a_{1n}^{(0)} & b_1^{(0)} \\ \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & a_{l,m+1}^{(0)} & \cdots & a_{lk}^{(0)} & \cdots & a_{ln}^{(0)} & b_l^{(0)} \\ \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & a_{m,m+1}^{(0)} & \cdots & a_{mk}^{(0)} & \cdots & a_{mn}^{(0)} & b_m^{(0)} \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} p_1 & \cdots & p_l & \cdots & p_m & p_{m+1} & \cdots & p_k & \cdots & p_n & b \\ 1 & \cdots & a_{1l}^{(1)} & \cdots & 0 & a_{1,m+1}^{(1)} & \cdots & 0 & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & a_{ll}^{(1)} & \cdots & 0 & a_{l,m+1}^{(1)} & \cdots & 1 & \cdots & a_{ln}^{(1)} & b_l^{(1)} \\ \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & a_{ml}^{(1)} & \cdots & 1 & a_{m,m+1}^{(1)} & \cdots & 0 & \cdots & a_{mn}^{(1)} & b_m^{(1)} \end{pmatrix},$$

换基：p_k 进入基底，p_l 换出基底；

⑦ 由基向量 $p_1, \cdots, p_{l-1}, p_k, p_{l+1}, \cdots, p_m$ 构造新的单纯形表，则可以得新的基本可行解 $x^{(1)}$；

⑧ 重复上述步骤直至满足所有的非基变量的检验数 $\sigma_j \leqslant 0$.

含最优解的单纯形表称为最终单纯形表（如表 10 – 4）

表 10 – 4

$c_j \to$			$c_1 \ \cdots$	c_l	$\cdots \ c_m \ \cdots$	c_j	$\cdots \ c_k \ \cdots$	θ
c_B	x_B	b	$x_1 \ \cdots$	x_l	$\cdots \ x_m \ \cdots$	x_j	$\cdots \ x_k \ \cdots$	
c_1	x_1	$b_1 - b_l \dfrac{a_{1k}}{a_{lk}}$	$1 \ \cdots$	$-\dfrac{a_{1k}}{a_{lk}}$	$\cdots \ 0 \ \cdots$	$a_{1j} - a_{1k}\dfrac{a_{lj}}{a_{lk}}$	$\cdots \ 0 \ \cdots$	θ_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_k	x_k	$\dfrac{b_l}{a_{lk}}$	$0 \ \cdots$	$\dfrac{1}{a_{lk}}$	$\cdots \ 0 \ \cdots$	$\dfrac{a_{lj}}{a_{lk}}$	$\cdots \ 1 \ \cdots$	θ_k
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_m	x_m	$b_m - b_l \dfrac{a_{mk}}{a_{lk}}$	$0 \ \cdots$	$-\dfrac{a_{mk}}{a_{lk}}$	$\cdots \ 1 \ \cdots$	$a_{mj} - a_{mk}\dfrac{a_{lj}}{a_{lk}}$	$\cdots \ 0 \ \cdots$	θ_m
σ_j			$0 \ \cdots$	$-\dfrac{c_k - z_k}{a_{lk}}$	$\cdots \ 0 \ \cdots$	$(c_j - z_j) - \dfrac{a_{lj}}{a_{lk}}(c_k - z_k)$	$\cdots \ 0 \ \cdots$	

例 10.4 用单纯形法求解线性规划问题

$$\max z = 3x_1 + x_2,$$

$$\text{s.t.} \begin{cases} x_1 + x_2 \leqslant 4, \\ -x_1 + x_2 \leqslant 2, \\ 6x_1 + 2x_2 \leqslant 18, \\ x_1, x_2 \geqslant 0. \end{cases}$$

解 先将问题化成标准形式

$$\max z = 3x_1 + x_2 + 0x_3 + 0x_4 + 0x_5,$$

$$\text{s.t.} \begin{cases} x_1 + x_2 + x_3 = 4, \\ -x_1 + x_2 + x_4 = 2, \\ 6x_1 + 2x_2 + x_5 = 18, \\ x_1, x_2, x_3, x_4, x_5 \geqslant 0. \end{cases}$$

单纯形法的迭代过程如表 10 – 5 所示.

表 10－5

$c_j \to$			3	1	0	0	0	θ
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5	
0	x_3	4	1	1	1	0	0	4
0	x_4	2	－1	1	0	1	0	－
0	x_5	18	[6]	2	0	0	1	3
σ_j			3	1	0	0	0	
0	x_3	5	0	$\frac{2}{3}$	1	0	$-\frac{1}{6}$	
0	x_4	1	0	$\frac{4}{3}$	0	1	$\frac{1}{6}$	
3	x_1	3	1	$\frac{1}{3}$	0	0	$\frac{1}{6}$	
σ_j			0	0	0	0	$-\frac{1}{2}$	

从表 10－5 中得出此问题的基可行解 $x=(3,0,5,1,0)^{\mathrm{T}}$ 为最优解，目标函数的最优值为 $z=9$.

10.3.2　用人工变量法找初始可行基——大 M 法和两段单纯形法

在前面的讨论中，将问题化为标准型后，如果系数矩阵中含有 $m\times m$ 阶单位矩阵，则可以找到一个初始可行基. 但在实际问题中，不一定都能直接找到初始可行基，这时可以采用人工变量法来确定初始可行基.

设线性规划问题的约束条件是 $\sum_{j=1}^{n} a_{ij}x_j = b_i (i=1,2,\cdots,m)$，分别给每个约束条件加入一个人工变量 $x_{n+1},\cdots,x_{n+m}$ 得

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + x_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + x_{n+2} = b_2, \\ \cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + x_{n+m} = b_m, \\ x_1, x_2, \cdots, x_{n+m} \geqslant 0. \end{cases}$$

这样以 $x_{n+1},\cdots,x_{n+m}$ 为基变量，可得到一个 $m\times m$ 阶单位矩阵. 令非基变量 $x_1,x_2,\cdots,x_n$ 为零，便得到一个初始基可行解 $x^{(0)}=(0,0,\cdots,0,b_1,b_2,\cdots,b_m)^{\mathrm{T}}$.

事实上,人工变量是加在原约束条件中的一个"虚拟变量",在求解过程中,经过基变换,可把这些人工变量从基变量中替换出去,使基变量中不再含有非零的人工变量,这时说明原问题有解. 若当所有 $\sigma_j \leqslant 0(j = m+1, m+2, \cdots, n)$ 时,在基变量中还含有非零的人工变量,则原问题无可行解. 现在的问题是:如何消除人工变量对目标函数的影响呢? 主要的方法有大 M 法和两阶段法.

1. 大 M 法

对于最大化问题,在约束条件中加入人工变量后,取人工变量在目标函数中的系数为 $(-M)$(M 为任意大的正数). 这样目标函数要实现最大化,必须把人工变量从基变量换出,否则目标函数不可能实现最大化. 相应地,对于最小化问题,则取人工变量在目标函数中的系数为 M. 下面举例说明:

例 10.5 用单纯形法求解线性规划问题

$$\max z = -3x_1 + x_3,$$

$$\text{s.t.}\begin{cases} x_1 + x_2 + x_3 \leqslant 4, \\ -2x_1 + x_2 - x_3 \geqslant 1, \\ 3x_2 + x_3 = 9, \\ x_1, x_2, x_3 \geqslant 0. \end{cases}$$

解 先把原问题化为标准型

$$\max z = -3x_1 + 0x_2 + x_3 + 0x_4 + 0x_5,$$

$$\text{s.t.}\begin{cases} x_1 + x_2 + x_3 + x_4 = 4, \\ -2x_1 + x_2 - x_3 - x_5 = 1, \\ 3x_2 + x_3 = 9, \\ x_1, x_2, x_3, x_4, x_5 \geqslant 0. \end{cases}$$

加入人工变量 x_6, x_7 得

$$\max z = -3x_1 + 0x_2 + x_3 + 0x_4 + 0x_5 - Mx_6 - Mx_7,$$

$$\text{s.t.}\begin{cases} x_1 + x_2 + x_3 + x_4 = 4, \\ -2x_1 + x_2 - x_3 - x_5 + x_6 = 1, \\ 3x_2 + x_3 + x_7 = 9, \\ x_1, x_2, x_3, x_4, x_5, x_6, x_7 \geqslant 0. \end{cases}$$

取 x_4,x_6,x_7 为基变量，令非基变量 x_1,x_2,x_3,x_5 为零，可以得到初始基可行解 $x^{(0)}=(0,0,0,4,0,1,9)^{\mathrm{T}}$. 列出初始单纯形表，用单纯形法进行求解(如表 10－6).

表 10－6

$c_j \to$			-3	0	1	0	0	$-M$	$-M$	θ
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
0	x_4	4	1	1	1	1	0	0	0	4
$-M$	x_6	1	-2	$[1]$	-1	0	-1	1	0	1
$-M$	x_7	9	0	3	1	0	0	0	1	3
σ_j			$-2M-3$	$4M$	1	0	$-M$	0	0	
0	x_4	3	3	0	2	1	1	-1	0	1
0	x_2	1	-2	1	-1	0	-1	1	0	—
$-M$	x_7	6	$[6]$	0	4	0	3	-3	1	1
σ_j			$6M-3$	0	$4M+1$	0	$-M$	0	0	
0	x_4	0	0	0	0	1	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	—
0	x_2	3	0	1	$\frac{1}{3}$	0	0	0	$\frac{1}{3}$	9
-3	x_1	1	1	0	$\left[\frac{2}{3}\right]$	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{6}$	$\frac{3}{2}$
σ_j			0	0	3	0	$\frac{2}{3}$	$-M-\frac{2}{3}$	$-M+\frac{1}{2}$	
0	x_4	0	0	0	0	1	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	
0	x_2	$\frac{5}{2}$	$-\frac{1}{2}$	1	0	0	$-\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	
1	x_3	$\frac{3}{2}$	$\frac{3}{2}$	0	1	0	$\frac{3}{4}$	$-\frac{3}{4}$	$\frac{1}{4}$	
σ_j			$-\frac{9}{2}$	0	3	0	$-\frac{3}{4}$	$-M-\frac{3}{4}$	$-M-\frac{1}{4}$	

于是可以得到问题的最优解为 $x=\left(0,\frac{5}{2},\frac{3}{2}\right)^{\mathrm{T}}$，最优值为 $z=\frac{3}{2}$.

2. 两阶段法

当用计算机求解含人工变量的线性规划问题时，如果试图使用大 M 法，因为每一台计算机都有一定字长的限制，于是只能用很大的数来代替充分大数 M，这样就可能造成计算上的错误. 为了避免此问题，可对添加人工变量后的线性规划问题分两个阶段来计算.

第一阶段：不考虑原规划是否存在基可行解，构造仅含人工变量的目标函数和实现最小化，即构造辅助线性规划

$$\min z = 0x_1 + 0x_2 + \cdots + 0x_n + x_{n+1} + \cdots + x_{n+m},$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + x_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + x_{n+2} = b_2, \\ \cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + x_{n+m} = b_m, \\ x_1, x_2, \cdots, x_{n+m} \geqslant 0. \end{cases}$$

容易看出，原问题有可行解（从而有基可行解）的充要条件是这个辅助的线性规划的最优值为零. 由于人工变量对应的列构成单位矩阵，故这个辅助线性规划必存在初始可行解. 若它的最优值等于零，即说明原问题存在基可行解. 此时可以进行第二阶段计算，否则原规划无可行解，停止计算.

第二阶段：由第一阶段得到原问题的一个基本可行解 $x^{(0)}$，把第一阶段的最终单纯形表中的最后一行及对应人工变量的列删去，作出原问题对应于 $x^{(0)}$ 的单纯形表，而将目标函数的系数行，换成原问题的目标函数系数，作为第二阶段计算的初始单纯形表.

下面对例 10.5 用两阶段法来求解.

解 第一阶段：作辅助线性规划问题

$$\min z = x_6 + x_7,$$

$$\text{s. t.} \begin{cases} x_1 + x_2 + x_3 + x_4 = 4, \\ -2x_1 + x_2 - x_3 - x_5 + x_6 = 1, \\ 3x_2 + x_3 + x_7 = 9, \\ x_1, x_2, x_3, x_4, x_5, x_6, x_7 \geqslant 0. \end{cases}$$

建立单纯形表（如表 10－7）：

表 10－7

$c_j\to$			0	0	0	0	0	－1	－1	θ
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
0	x_4	4	1	1	1	1	0	0	0	4
－1	x_6	1	－2	[1]	－1	0	－1	1	0	1
－1	x_7	9	0	3	1	0	0	0	1	3
σ_j			－2	4	0	0	－1	0	0	
0	x_4	3	3	0	2	1	1	－1	0	1
0	x_2	1	－2	1	－1	0	－1	1	0	－
－1	x_7	6	[6]	0	4	0	3	－3	1	1
σ_j			6	0	4	0	3	－4	0	
0	x_4	0	0	0	0	1	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	
0	x_2	3	0	1	$\frac{1}{3}$	0	0	0	$\frac{1}{3}$	
0	x_1	1	1	0	$\frac{2}{3}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{6}$	
σ_j			0	0	0	0	0	－1	－1	

至此，所有非基变量的检验数都小于 0，则已得到最优解，且最优值为 0，故已求得原问题一个基可行解.

第二阶段：现转入求解原问题的解，建立初始单纯形表如表 10－8 所示.

表 10－8

c_j			－3	0	1	0	0	θ
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5	
0	x_4	0	0	0	0	1	$-\frac{1}{2}$	－
0	x_2	3	0	1	$\frac{1}{3}$	0	0	9
－3	x_1	1	1	0	$\left[\frac{2}{3}\right]$	0	$\frac{1}{2}$	$\frac{3}{2}$

续表

c_j			-3	0	1	0	0	θ
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5	
σ_j			0	0	3	0	$\frac{3}{2}$	
0	x_4	0	0	0	0	1	$-\frac{1}{2}$	
0	x_2	$\frac{5}{2}$	$-\frac{1}{2}$	1	0	0	$-\frac{1}{4}$	
1	x_3	$\frac{3}{2}$	$\frac{3}{2}$	0	1	0	$\frac{3}{4}$	
σ_j			$-\frac{9}{2}$	0	0	0	$-\frac{3}{4}$	

故原问题的最优化解为 $x=\left(0,\frac{5}{2},\frac{3}{2}\right)^{\mathrm{T}}$,最优值为 $z=\frac{3}{2}$.

10.3.3 修正单纯形法

下面对单纯形法的计算过程用矩阵来描述,以便加强对单纯形法的理解及改进.

对于线性规划问题的标准型

$$\max z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$
$$\text{s. t.} \begin{cases} \boldsymbol{Ax} = \boldsymbol{b}, \\ \boldsymbol{x} \geqslant 0. \end{cases}$$

不妨设 $\boldsymbol{B}$ 是一个可行基,于是系数矩阵 $\boldsymbol{A}$ 可分块为 $\boldsymbol{A}=(\boldsymbol{B},\boldsymbol{N})$,不失一般性,假定 $\boldsymbol{B}=(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_m)$,对应于 $\boldsymbol{B}$ 的基变量为 $\boldsymbol{x}_B=(x_1,\cdots,x_m)^{\mathrm{T}}$,$\boldsymbol{N}=(\boldsymbol{p}_{m+1},\cdots,\boldsymbol{p}_n)$对应的非基变量为 $\boldsymbol{x}_N=(x_{m+1},\cdots,x_n)^{\mathrm{T}}$,则 $\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_B\\ \boldsymbol{x}_N\end{bmatrix}$,相应地 $\boldsymbol{c}^{\mathrm{T}}=(\boldsymbol{c}_B,\boldsymbol{c}_N)$,其中 $\boldsymbol{c}_B$ 为基变量 $\boldsymbol{x}_B$ 的系数行向量,$\boldsymbol{c}_N$ 是非基变量 $\boldsymbol{x}_N$ 的系数行向量,于是原问题化为

$$\max z = \boldsymbol{c}_B\boldsymbol{x}_B + \boldsymbol{c}_N\boldsymbol{x}_N, \tag{10-12}$$
$$\text{s. t.} \begin{cases} \boldsymbol{Bx}_B + \boldsymbol{Nx}_N = \boldsymbol{b}, & (10-13) \\ \boldsymbol{x}_B,\boldsymbol{x}_N \geqslant 0. & (10-14) \end{cases}$$

则

$$x_B = B^{-1}b - B^{-1}Nx_N. \tag{10-15}$$

将(10-15)式代入目标函数(10-12),得

$$z = c_BB^{-1}b + (c_N - c_BB^{-1}N)x_N \tag{10-16}$$

令 $x_N=0$ 得 $x_B=B^{-1}b$,从而得到相应的基可行解和目标函数值分别为

$$x = \begin{pmatrix} x_B \\ x_N \end{pmatrix} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix} \text{和} z = c_BB^{-1}b.$$

从(10-15)、(10-16)式中可以看出,非基变量的系数($c_N - c_BB^{-1}N$)就是非基变量的检验数 $\sigma_j(j=m+1,\cdots,n)$.

将(10-15)、(10-16)式改写为如下形式

$$\begin{cases} x_B + B^{-1}Nx_N = B^{-1}b, \\ z - (c_N - c_BB^{-1}N)x_N = c_BB^{-1}b. \end{cases}$$

令 $\sigma_N = c_N - c_BB^{-1}N, \sigma = c - c_BB^{-1}A$,则

$$z = c_BB^{-1}b + \sigma_Nx_N = c_BB^{-1}b + \sigma x, \text{其中} \sigma = (0, \sigma_N).$$

熟悉了单纯形法的表格计算方法和矩阵形式,在迭代过程中做了很多与下一步迭代无关的重复计算,影响了计算效率,用计算机编程求解时,既占用内存单元,同时会影响计算的精度.通过分析可以看出,在整个迭代过程中,基矩阵的逆矩阵 B^{-1} 的求解是关键.只要求出 B^{-1},则单纯形表中其他行和列的数字也随之确定了,故提出了修正单纯形法.

下面是修正单纯形法的算法:

算法 10.2

(一) 初始步

做出初始单纯形表

由 $Ax=b$,其中 $A=(a_{ij})_{m\times n}, x\in \mathbf{R}^n, b\in \mathbf{R}^m$.将 $A=(B,N), x=\begin{bmatrix} x_B \\ x_N \end{bmatrix}$代入变为

$$Bx_N + Nx_N = b,$$

由此得到一个初始的基本可行解

$$x^{(0)} = \begin{pmatrix} x_B \\ x_N \end{pmatrix} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}.$$

(二) 迭代步

① 给出初始基本可行解 $x^{(0)}$;

② 计算非基变量的检验数 $\sigma_N = c_N - c_BB^{-1}N$;

若 $\sigma_N \leqslant 0, x^{(0)} \Rightarrow x^*$,计算结束;

否则转③.

③ 令 $\sigma_k=\max(\sigma_j), j\in \mathbf{N}$，计算 $\boldsymbol{B}^{-1}\boldsymbol{p}_k$；

④ 检查 $\boldsymbol{B}^{-1}\boldsymbol{p}_k\leqslant 0$ 是否成立？

是，则原问题无解，停止计算，

否，转⑤；

⑤ 计算 $\theta^{(0)}=\min\left\{\frac{(\boldsymbol{B}^{-1}\boldsymbol{b})_i}{(\boldsymbol{B}^{-1}\boldsymbol{p}_k)_i}\middle|(\boldsymbol{B}^{-1}\boldsymbol{b})_i>0\right\}=\frac{(\boldsymbol{B}^{-1}\boldsymbol{b})_l}{(\boldsymbol{B}^{-1}\boldsymbol{p}_k)_l}$，由此定出主列为第 k 列，主行为第 l 行，得到新的基变量和基矩阵 $\boldsymbol{B}_1$；

⑥ 计算新的基矩阵的逆矩阵 $\boldsymbol{B}_1^{-1}$，求出 $\boldsymbol{B}_1^{-1}\boldsymbol{b}$ 及 $\boldsymbol{c}_B\boldsymbol{B}_1^{-1}$；

⑦ 重复上述步骤直至满足 $\sigma_N\leqslant 0$.

在初始单纯形表中，由于 $\boldsymbol{B}$ 是单位矩阵，故 $\boldsymbol{B}^{-1}$ 也是单位矩阵，所以修正单纯形在开始计算时，不需要计算基的逆矩阵，但经过一次迭代后，需要计算新的基矩阵的逆矩阵 $\boldsymbol{B}_1^{-1}$，而 $\boldsymbol{B}_1^{-1}$ 的求解比较繁琐. 但注意到上一步迭代的基 $\boldsymbol{B}$ 与下一步迭代的基 $\boldsymbol{B}_1$ 之间只相差一个列向量. 故可用如下简单算法：

设 $\boldsymbol{B}_1^{-1}=\boldsymbol{E}\boldsymbol{B}^{-1}$，其中 $\boldsymbol{E}=(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_{l-1},\boldsymbol{\xi},\boldsymbol{e}_{l-1},\cdots,\boldsymbol{e}_m)$，$\boldsymbol{e}_i$ 表示第 i 个位置的元素为 1，其他元素为 0 的单位列向量.

$$\boldsymbol{\xi}=\begin{pmatrix}-a_{1k}/a_{lk}\\ -a_{2k}/a_{lk}\\ \vdots\\ 1/a_{lk}\\ \vdots\\ -a_{mk}/a_{lk}\end{pmatrix}\begin{matrix}\\ \\ \\ \leftarrow \text{第 } l \text{ 行}\\ \\ \\ \end{matrix}$$

下面举例说明修正单纯形法的计算步骤.

例 10.6　用修正单纯形法求解线性规划问题

$$\max z=2x_1+3x_2,$$

$$\text{s.t.}\begin{cases}x_1+2x_2\leqslant 8,\\ x_1\leqslant 4,\\ x_2\leqslant 3,\\ x_1,x_2,x_3\geqslant 0.\end{cases}$$

解　化为标准型

$$\max z = 2x_1 + 3x_2 + 0x_3 + 0x_4 + 0x_5,$$

$$\text{s. t.}\begin{cases} x_1 + 2x_2 + x_3 = 8, \\ x_1 + x_4 = 4, \\ x_2 + x_5 = 3, \\ x_1, x_2, x_3, x_4, x_5 \geqslant 0. \end{cases}$$

其中

$$\boldsymbol{A} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_5) = \begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 8 \\ 4 \\ 3 \end{pmatrix},$$

$$\boldsymbol{c}^{\mathrm{T}} = (c_1, c_2, c_3, c_4, c_5) = (2,3,0,0,0).$$

取初始基 $\boldsymbol{B}_0 = (\boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_5) = \boldsymbol{I}$，则 $\boldsymbol{B}_0^{-1} = \boldsymbol{B}_0 = \boldsymbol{I}, \boldsymbol{B}_0^{-1}\boldsymbol{b} = (8,4,3)^{\mathrm{T}}$，$\boldsymbol{x}_{B_0} = (x_3, x_4, x_5)^{\mathrm{T}}, \boldsymbol{c}_{B_0} = (0,0,0), \boldsymbol{x}_{N_0} = (x_1, x_2)^{\mathrm{T}}, \boldsymbol{c}_{N_0} = (2,3)$.

计算非基变量检验数 $\boldsymbol{\sigma}_{N_0} = \boldsymbol{c}_{N_0} - \boldsymbol{c}_{B_0}\boldsymbol{B}_0^{-1}\boldsymbol{N}_0 = (2,3)$，由此确定 x_2 为换入变量，计算

$$\theta = \min\left\{\frac{(\boldsymbol{B}_0^{-1}\boldsymbol{b})_i}{(\boldsymbol{B}_0^{-1}\boldsymbol{p}_2)_i} \middle| \boldsymbol{B}_0^{-1}\boldsymbol{p}_2 > 0\right\} = \min\left\{\frac{8}{2}, -, 3\right\} = 3.$$

对应的换出变量为 x_5，故得到新的基 $\boldsymbol{B}_1 = (\boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_2), \boldsymbol{x}_{B_1} = (x_3, x_4, x_2)^{\mathrm{T}}$，$\boldsymbol{c}_{B_1} = (0,0,3), \boldsymbol{c}_{N_1} = (2,0)$.

第 1 次迭代，计算

$$\boldsymbol{\xi}_1 = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{B}_1^{-1} = \boldsymbol{E}_1\boldsymbol{B}_0^{-1} = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

计算非基变量检验数

$$\begin{aligned} \boldsymbol{\sigma}_{N_1} &= \boldsymbol{c}_{N_1} - \boldsymbol{c}_{B_1}\boldsymbol{B}_1^{-1}\boldsymbol{N}_1 \\ &= (2,0) - (0,0,3)\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= (2, -3), \end{aligned}$$

由此确定 x_1 为换入变量，计算

$$\theta = \min\left\{\frac{(B_1^{-1}b)_i}{(B_1^{-1}p_1)_i} \middle| B_1^{-1}p_1 > 0\right\} = \min\left\{\frac{2}{1}, \frac{4}{1}, -\right\} = 2,$$

对应的换出变量为 x_3，故得到新的基 $B_2 = (p_1, p_4, p_2)$，$x_{B_2} = (x_1, x_4, x_2)^T$，$c_{B_2} = (2,0,3)$，$c_{N_2} = (0,0)$.

第 2 次迭代，计算

$$\xi_2 = \begin{pmatrix} \frac{1}{1} \\ -\frac{1}{1} \\ 0 \end{pmatrix},$$

$$B_2^{-1} = E_2 B_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -2 \\ -1 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

计算非基变量检验数

$$\begin{aligned} \sigma_{N_2} &= c_{N_2} - c_{B_2} B_2^{-1} N_2 \\ &= (0,0) - (2,0,3)\begin{pmatrix} 1 & 0 & -2 \\ -1 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \\ &= (-2,1), \end{aligned}$$

由此确定 x_5 为换入变量，计算

$$\theta = \min\left\{\frac{(B_2^{-1}b)_i}{(B_2^{-1}p_5)_i} \middle| B_2^{-1}p_5 > 0\right\} = \min\left\{-, \frac{2}{2}, \frac{3}{1}\right\} = 1,$$

对应的换出变量为 x_4，故得到新的基 $B_3 = (p_1, p_5, p_2)$，$x_{B_3} = (x_1, x_5, x_2)^T$，$c_{B_3} = (2,0,3)$，$c_{N_2} = (0,0)$.

第 3 次迭代，可得最优解为 $x = (4,2)^T$，最优值为 $z = 14$.

10.4 线性规划的对偶问题

10.4.1 对偶问题举例

例 10.7 某电器元件生产厂家在计划期内要安排生产甲、乙两种产品，已知生产单位产品所需的设备台时及 A、B 两种原材料的消费如表 10-9 所示. 该工厂每生产一件产品甲可获利 1 元，每生产一件产品乙可获利 2 元，问应如何安排生产计划使该厂获得利润最大?

表 10－9

	甲	乙	限量
设备	5	0	15 台时
原材料 A	2	6	24 kg
原材料 B	1	1	5 kg

解　设 x_1,x_2 分别表示在计划期内产品甲和乙的产量，则有线性规划问题

$$\max z = x_1 + 2x_2,$$

$$\text{s.t.}\begin{cases}5x_1 \leqslant 15,\\ 2x_1 + 6x_2 \leqslant 24,\\ x_1 + x_2 \leqslant 5,\\ x_1,x_2 \geqslant 0.\end{cases}$$

现在的问题是：若由工厂的决策者决定不生产甲、乙两种产品，而将其所有资源和产品的生产权出售，那么该如何给每种资源制定价格，使出售后所得利润不低于该厂自己生产所得利润？

设用 y_1,y_2,y_3 分别表示出租单位设备台时的租金和出让单位原材料 A、B 的附加额. 决策者在制定决策时，首先会作如下的比较，若用 5 个单位设备台时、2 个单位原材料 A 和 1 单位原材料 B 可以生产一件产品甲，可获利 1 元，那么生产每件产品的设备台时的出租费和原材料 A、B 的出让附加额应不低于生产 1 件产品甲的利润，即

$$5y_1 + 2y_2 + y_3 \geqslant 1,$$

同理，生产每件产品 B 的设备台时的出租费和原材料的出让附加额也应不低于生产 1 件产品 B 的利润，即

$$6y_2 + y_3 \geqslant 2.$$

把工厂所有设备台时出租和资源全部出让，其收入为

$$w = 15y_1 + 24y_2 + 5y_3.$$

从工厂的决策者来看，当然是 w 越大越好；但从接收者来看，他的支付金额越少越好. 所以工厂的决策只能在满足不少于自己生产产品的利润条件下，使其总收入尽可能地大. 故须求解如下的问题.

$$\min w = 15y_1 + 24y_2 + 5y_3,$$

$$\text{s.t.}\begin{cases}5y_1+2y_2+y_3\geqslant 1,\\ 6y_2+y_3\geqslant 2,\\ y_1,y_2,y_3\geqslant 0.\end{cases}$$

这个问题就是原线性规划问题的对偶问题,它也是一个线性规划问题.

10.4.2 原问题与对偶问题的关系

线性规划的原问题和对偶问题的一般形式有两种:

(1)对称形式的对偶问题

设原问题为

$$\max z=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$$

$$\text{s.t.}\begin{cases}\boldsymbol{Ax}\leqslant\boldsymbol{b},\\ \boldsymbol{x}\geqslant 0.\end{cases}$$

则其对偶问题为

$$\min z'=\boldsymbol{yb},$$

$$\text{s.t.}\begin{cases}\boldsymbol{yA}\geqslant\boldsymbol{c}^{\mathrm{T}},\\ \boldsymbol{y}\geqslant 0.\end{cases}$$

将这两个对称形式的问题进行比较,可以得出它们之间的对应关系:

① 原问题中的约束条件个数等于它的对偶问题中的变量个数;

② 原问题的目标函数的系数是它的对偶问题中的约束条件的右端项;

③ 原问题的目标函数为最大化,则它的对偶问题目标函数为最小化;

④ 原问题的约束条件为“$\leqslant$”,它的对偶问题的约束条件为“$\geqslant$”.

(2)非对称形式的对偶问题.

设原问题为

$$\max z=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

$$\text{s.t.}\begin{cases}\boldsymbol{Ax}=\boldsymbol{b},\\ \boldsymbol{x}\geqslant 0.\end{cases}$$

原问题中的约束条件为等式,对偶问题应该如何表达呢?

由于 $\boldsymbol{Ax}=\boldsymbol{b}$ 等价于$\begin{cases}\boldsymbol{Ax}\leqslant\boldsymbol{b},\\ -\boldsymbol{Ax}\leqslant-\boldsymbol{b},\end{cases}$故上式可改写为

$$\max z = c^{\mathrm{T}} x,$$

$$\text{s.t.} \begin{cases} Ax \leqslant b \cdots\cdots\cdots\cdots\cdots u, \\ -Ax \leqslant -b \cdots\cdots\cdots v, \\ x \geqslant 0. \end{cases}$$

按对称形式写出它的对偶问题

$$\min z' = ub - vb,$$

$$\text{s.t.} \begin{cases} uA - vA \geqslant c^{\mathrm{T}}, \\ u \geqslant 0, v \geqslant 0. \end{cases}$$

即

$$\min z' = (u - v)b,$$

$$\text{s.t.} \begin{cases} (u - v)A \geqslant c^{\mathrm{T}}, \\ u \geqslant 0, v \geqslant 0, \end{cases}$$

记 $y = u - v$,则

$$\min z' = yb,$$

$$\text{s.t.} \begin{cases} yA \geqslant c^{\mathrm{T}}, \\ y \text{ 无非负约束}. \end{cases}$$

通过上面两种线性规划问题与对偶问题的比较可知,区别在于约束条件的形式与其对应变量的取值. 一般说来,对称或非对称线性规划原问题与对偶问题的对应关系可归纳为表 10 - 10 所示.

表 10 - 10

原问题(或对偶问题)	对偶问题(或原问题)
目标函数　$\max z$	目标函数　$\min w$
变量 n 个	约束条件 n 个
变量 $\geqslant 0$	约束条件 $\geqslant 0$
变量 $\leqslant 0$	约束条件 $\leqslant 0$
变量无约束	约束条件 =
约束条件 m 个	变量 m 个
约束条件 $\leqslant 0$	变量 $\geqslant 0$
约束条件 $\geqslant 0$	变量 $\leqslant 0$
约束条件 = 0	变量无约束
约束条件右端项	目标函数变量的系数
目标函数变量的系数	约束条件右端项

例 10.8 试求下列问题的对偶问题

$$\max z = 10y_1 + 8y_2 + y_3,$$

$$\text{s.t.}\begin{cases} y_1 + 2y_2 \geqslant 3, \\ y_1 + y_3 \leqslant 2, \\ -3y_1 + 2y_2 + y_3 \leqslant -4, \\ y_1 - y_2 + y_3 = 1, \\ y_1 \geqslant 0, y_2 \leqslant 0, y_3 \text{ 无约束}. \end{cases}$$

解 原问题的对偶问题为

$$\min w = 3x_1 + 2x_2 - 4x_3 + x_4,$$

$$\text{s.t.}\begin{cases} x_1 + x_2 - 3x_3 + x_4 \geqslant 10, \\ 2x_1 + 2x_3 - x_4 \leqslant 8, \\ x_2 + x_3 + x_4 = 1, \\ x_1 \leqslant 0, x_2, x_3 \geqslant 0, x_4 \text{ 无约束}. \end{cases}$$

10.4.3 对偶问题的基本定理

定理 10.4 对偶问题的对偶问题是原问题.

证 设原问题为

$$\max z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

$$\text{s.t.}\begin{cases} \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}, \\ \boldsymbol{x} \geqslant 0, \end{cases}$$

则其对偶问题为

$$\min w = \boldsymbol{y}\boldsymbol{b},$$

$$\text{s.t.}\begin{cases} \boldsymbol{y}\boldsymbol{A} \geqslant \boldsymbol{c}^{\mathrm{T}}, \\ \boldsymbol{y} \geqslant 0. \end{cases}$$

两边同乘以 -1 得

$$-\min w = -\boldsymbol{y}\boldsymbol{b},$$

$$\text{s.t.}\begin{cases}-\boldsymbol{y}\boldsymbol{A} \leqslant -\boldsymbol{c}^{\mathrm{T}},\\ \boldsymbol{y} \geqslant 0.\end{cases}$$

因 $\min(w) = \max(-w)$，所以

$$\max(-w) = -\boldsymbol{y}\boldsymbol{b},$$

$$\text{s.t.}\begin{cases}-\boldsymbol{y}\boldsymbol{A} \leqslant -\boldsymbol{c}^{\mathrm{T}},\\ \boldsymbol{y} \geqslant 0.\end{cases}$$

此式对偶问题为

$$\min(-w') = -\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

$$\text{s.t.}\begin{cases}-\boldsymbol{A}\boldsymbol{x} \geqslant -\boldsymbol{b},\\ \boldsymbol{x} \geqslant 0.\end{cases}$$

又 $\min(-w') = \max(w')$，所以

$$\max w' = \max z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

$$\text{s.t.}\begin{cases}\boldsymbol{A}\boldsymbol{x} \leqslant b,\\ \boldsymbol{x} \geqslant 0.\end{cases}$$

故结论成立.

定理 10.5（弱对偶定理）　若 $\bar{\boldsymbol{x}}$ 和 $\bar{\boldsymbol{y}}$ 分别为原问题和对偶问题规划的可行解，则存在 $\boldsymbol{c}^{\mathrm{T}}\bar{\boldsymbol{x}} \leqslant \bar{\boldsymbol{y}}\boldsymbol{b}$.

证　设原问题为

$$\max z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

$$\text{s.t.}\begin{cases}\boldsymbol{A}\bar{\boldsymbol{x}} \leqslant \boldsymbol{b},\\ \boldsymbol{x} \geqslant 0.\end{cases}$$

因 $\bar{\boldsymbol{x}}$ 为原问题的可行解，故 $\boldsymbol{A}\bar{\boldsymbol{x}} \leqslant \boldsymbol{b}$.

若 $\bar{\boldsymbol{y}}$ 是给定的一组值，设它是对偶问题的可行解，则将 $\bar{\boldsymbol{y}}$ 左乘上式得 $\bar{\boldsymbol{y}}\boldsymbol{A}\bar{\boldsymbol{x}} \leqslant \bar{\boldsymbol{y}}\boldsymbol{b}$.

原问题的对偶问题为

$$\min w = \boldsymbol{y}\boldsymbol{b},$$

$$\text{s.t.}\begin{cases} yA \geqslant c^{T}, \\ y \geqslant 0. \end{cases}$$

因 $\bar{y}$ 为对偶问题的可行解，故 $\bar{y}A \geqslant c^{T}$. 将 $\bar{x}$ 右乘上式得 $\bar{y}A\bar{x} \geqslant c^{T}\bar{x}$. 故

$$\bar{y}b \geqslant \bar{y}A\bar{x} \geqslant c^{T}\bar{x},$$

则结论得证.

由此定理可推出以下结论：

推论 10.1 若 $\bar{x}$ 为问题的任一可行解，则 $c\bar{x}$ 为其对偶问题目标函数值的一个下界；若 $\bar{y}$ 为对偶问题的任一可行解，则 $\bar{y}b$ 为原问题目标函数值的一个上界.

推论 10.2 若原问题有可行解，但其目标函数值无上界，则其对偶问题无可行解；若对偶问题有可行解，但其目标函数值无下界，则其原问题无可行解.

推论 10.3 若原问题有可行解，但其对偶问题无可行解，则原问题无上界；若对偶问题有可行解，但其原问题无可行解，则对偶问题无下界.

例 10.9 试说明线性规划问题

$$\max z_1 = x_1 + x_2,$$

$$\text{s.t.}\begin{cases} -x_1 + x_2 + x_3 \leqslant 2, \\ -2x_1 + x_2 - x_3 \leqslant 1, \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

为无界解.

解 原问题的对偶问题为

$$\min z_2 = 2y_1 + y_2,$$

$$\text{s.t.}\begin{cases} -y_1 - 2y_2 \geqslant 1, \\ y_1 + y_2 \geqslant 1, \\ y_1 - y_2 \geqslant 0, \\ y_1, y_2 \geqslant 0. \end{cases}$$

由于原问题有可行解 $x_1 = x_2 = x_3 = 0$，但对偶问题无可行解（第一个约束条件不可能），所以由推论 10.3 可知，原问题无上界，即目标函数可任意大.

定理 10.6　若原问题(对偶问题)为无界解,则其对偶问题(原问题)无可行解.

证　由弱对偶性可得. 但注意,定理的逆不成立,即当原问题(对偶问题)无可行解时,其对偶问题(原问题)或具有无界解或无可行解.

定理 10.7　设 x^* 是原问题可行解,y^* 是对偶问题可行解,且 $c^Tx^*=y^*b$,则 x^*,y^* 是原问题和对偶问题的最优解.

证　设 $\bar{x}$ 为原问题任一可行解,$c^Tx^*=y^*b\geqslant c^T\bar{x}$,故 x^* 为原问题的最优解,同理可证,y^* 是对偶问题的最优解.

定理 10.8(对偶问题)　若原问题有最优解,那么对偶问题也有最优解,其最优值相等.

证　设 x^* 为原问题最优解,它对应不变基矩阵 B 必存在,则有 $c^T-c_BB^{-1}A\leqslant 0$,即得 $y^*A\geqslant c^T$,其中 $y^*=c_BB^{-1}$. 若 y^* 是对偶问题的可行解,它使 $w=y^*b=c_Bb^{-1}b$,因原问题的最优解是 x^*,使目标函数取值 $z=c^Tx^*=c_BB^{-1}b$. 由此得 $y^*b=c_BB^{-1}b=c^Tx^*$,即 y^* 为最优解.

定理 10.9(互补松弛定理)　若 x^*,y^* 分别是原问题和对偶问题的可行解,那么 $y^*x_S=0$ 和 $y_Sx^*=0$,当且仅当 x^*,y^* 分别为原问题和对偶问题的最优解.

证　必要性: 设原问题为

$$\max z=c^Tx,$$

$$\text{s. t.}\begin{cases}Ax+x_S=b,\\ x,x_S\geqslant 0.\end{cases}$$

其对偶问题为

$$\min w=yb,$$

$$\text{s. t.}\begin{cases}yA-y_S=c,\\ y,y_S\geqslant 0.\end{cases}$$

将 $c^T=yA-y_S$ 代替 c^T 得

$$z=(yA-y_S)x=yAx-y_Sx. \tag{10-17}$$

用 $b=Ax+x$ 代替 b 得

$$w=y(Ax+x_S)=yAx+yx_S. \tag{10-18}$$

若 $y_Sx^*=0$,$y^*x_S=0$,则 $y^*b=y^*Ax^*=c^Tx^*$,由定理 10.8 知,x^*、

y^* 为最优解.

充分性：若 x^*、y^* 分别为原问题与对偶问题的最优解，由定理 10.7 知

$$y^* b = y^* A x^* = c^T x^*.$$

根据(10－17)、(10－18)两式，可得 $y_S x^* = 0, y^* x_S = 0$.

例 10.10 设线性规划

$$\min w = 2x_1 + 3x_2 + 5x_3 + 2x_4 + 3x_5,$$

$$\text{s.t.} \begin{cases} x_1 + x_2 + 2x_3 + x_4 + 3x_5 \geqslant 4, \\ 2x_1 - x_2 + 3x_3 + x_4 + x_5 \geqslant 3, \\ x_1, x_2, x_3, x_4, x_5 \geqslant 0. \end{cases}$$

的对偶问题的最优解为 $y_1^* = \dfrac{4}{5}, y_2^* = \dfrac{3}{5}, z = 5$，找出原问题的最优解.

解 问题的对偶问题为

$$\max z = 4y_1 + 3y_2,$$

$$\text{s.t.} \begin{cases} y_1 + 2y_2 \leqslant 2, & (10-19) \\ y_1 - y_2 \leqslant 3, & (10-20) \\ 2y_1 + 3y_2 \leqslant 5, & (10-21) \\ y_1 + y_2 \leqslant 2, & (10-22) \\ 3y_1 + y_2 \leqslant 3, & (10-23) \\ y_1, y_2 \geqslant 0. \end{cases}$$

将 y_1^*, y_2^* 代入约束条件得(10－20)，(10－21)，(10－22)为严格不等式. 由互补松弛性得 $x_2^* = x_3^* = x_4^* = 0$.

因 $y_1, y_2 \geqslant 0$，原问题的约束条件应取等式，故得

$$\begin{cases} x_1^* + 3x_5^* = 4, \\ 2x_1^* + x_5^* = 3. \end{cases}$$

求解得 $x_1^* = 1, x_5^* = 1$. 故原问题的最优解为 $x^* = (1,0,0,0,1)^T$.

10.4.4 对偶单纯形法

由对偶的自反性，通过求对偶问题的解来实现求解原问题解的目的，

这种方法称为对偶单纯形法. 从对偶问题的一个可行解出发,在保持对对偶问题可行的前提下,对原问题的非可行解进行迭代,使其逐步向可行解"靠近",当它最终变为原问题的可行解时,也就同时得到了原问题及对偶问题的最优解. 具体求解算法如下:

算法 10.3

(一) 初始步

① 建立初始单纯形表.

(二) 迭代步

② 进行最优性检验,若现行常数列 $\boldsymbol{b}\geqslant 0$,检验数都为非正,则现行对偶问题的基可行解即为最优解,停止计算. 若常数列 $\boldsymbol{b}$ 至少还有一个负分量,检验数保持非正,则转③;

③ 确定换出变量,将现行 $\boldsymbol{b}$ 中最小的负元素所在行,即

$$\min\{(\boldsymbol{B}^{-1}\boldsymbol{b})_i \mid (\boldsymbol{B}^{-1}\boldsymbol{b})_i < 0\} = (\boldsymbol{B}^{-1}\boldsymbol{b})_l,$$

对应的基变量 x_l 作为换出变量;

④ 确定换入变量. 检查 x_l 所在行的各系数 $a_{lj}(j=1,2,\cdots,n)$. 若所有 $a_{lj}\geqslant 0$,则无可行解,停止计算. 若存在 $a_{lj}<0\ \ (j=1,2,\cdots,n)$,计算

$$\theta = \min_j\left\{\frac{\sigma_j}{a_{lj}}\ \middle|\ a_{lj}<0\right\} = \frac{\sigma_k}{a_{lk}},$$

则确定 x_k 为换入变量;

⑤ 以 a_{lk} 为主元素,按原单纯形法进行迭代运算;

⑥ 重复②—⑤的步骤.

例 10.11　用对偶单纯形法求解

$$\min w = 4x_1 + 3x_2 + 8x_3,$$

$$\text{s.t.}\begin{cases} x_1 + x_3 \geqslant 2, \\ x_2 + 2x_3 \geqslant 5, \\ x_j \geqslant 0 \quad (j=1,2,3). \end{cases}$$

解　先将问题改写为

$$\max z = -4x_1 - 3x_2 - 8x_3,$$

$$\text{s.t.}\begin{cases} -x_1 - x_3 \leqslant -2, \\ -x_2 - 2x_3 \leqslant -5, \\ x_j \geqslant 0 \quad (j=1,2,3). \end{cases}$$

再化为标准型

$$\max z = -4x_1 - 3x_2 - 8x_3 + 0x_4 + 0x_5,$$

$$\text{s.t.} \begin{cases} -x_1 - x_3 + x_4 = -2, \\ -x_2 - 2x_3 + x_5 = -5, \\ x_j \geqslant 0 \quad (j = 1,2,3,4,5). \end{cases}$$

建立初始单纯形表(见表 10－11)

表 10－11

c_j			-4	-3	-8	0	0
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5
0	x_4	-2	-1	0	-1	1	0
0	x_5	-5	0	[-1]	-2	0	1
σ_j			-4	-3	-8	0	0

由算法 10.3 的②,计算 $\min\{-2,-5\} = -5$,故 x_5 为换出变量;

由算法 10.3 的③,计算 $\theta = \min\left\{\dfrac{-3}{-1},\dfrac{-8}{-2}\right\} = 3$;故 x_2 为换入变量.

按计算步骤进行迭代,得到表 10－12

表 10－12

c_j			-4	-3	-8	0	0
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5
0	x_4	-2	-1	0	-1	1	0
-3	x_2	5	0	1	2	0	-1
σ_j			-4	0	-2	0	-3

由算法 10.3 的②和③,x_4 为换出变量,故 x_3 为换入变量,得表 10－13.

表 10－13

c_j			-4	-3	-8	0	0
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5
-8	x_3	2	1	0	1	-1	0
-3	x_2	1	-2	1	0	2	-1
σ_j			-2	0	0	-2	-3

所以最优解为 $x = (0,1,2)^T$,最优值为 $z = 19$.

第 11 章　整数规划

如果一个数学规划的某些决策变量或全部决策变量要求必须取整数，则这样的问题称为整数规划问题，其模型称为整数规划模型（integer programming，简记 IP）.

在整数规划中，如果要求全部决策变量都取整数值，则称此规划为纯整数规划或全整数规划；如果仅有一部分决策变量要求取整数，则称此规划为混合整数规划；特别地，如果所有的决策变量的取值仅限于 0 或 1，就称此规划为 0－1 整数规划，即是整数规划的一种特殊情形. 如果整数规划的目标函数和约束条件都是线性的，则称此问题为整数线性规划问题. 在这里只就整数线性规划问题进行讨论.

11.1　整数规划问题举例

例 11.1　下料问题

设用某种型号的钢材下 m 种不同的零件毛坯 $A_i(i=1,2,\cdots,m)$，每种零件毛坯的需求量是 $a_i(i=1,2,\cdots,m)$，现有 n 种不同的下料方式 B_j $(j=1,2,\cdots,n)$，设按 B_j 方式下料 A_i 的件数设为 $c_{ij}(i=1,\cdots,m,j=1,\cdots,n)$（如表 11－1 所示）. 问应怎样安排下料方式，使既满足需求，又使所用原材料最省？

表 11－1

	B_1	B_2，	$\cdots$，	B_n	各零件需求量
A_1	c_{11}	c_{12}	$\cdots$	c_{1n}	a_1
A_2	c_{21}	c_{22}	$\cdots$	c_{2n}	a_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_m	c_{m1}	c_{m2}	$\cdots$	c_{mn}	a_m

解　设决策变量 x_j 表示用 B_j 方式下料的原材料数，则目标函数为 n 种方式下料的原材料总数为

$$z=\sum_{j=1}^{n}x_j,$$

约束条件为

$$\sum_{j=1}^{n} c_{ij}x_j \geqslant a_i \quad (i=1,2,\cdots,m),$$

$$x_j \geqslant 0, \text{且为整数},$$

该问题的数学模型为

$$\min z = \sum_{j=1}^{n} x_j,$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{n} c_{ij}x_j \geqslant a_i, i=1,2,\cdots,m, \\ x_j \geqslant 0, \text{且为整数}. \end{cases}$$

这是一个纯整数规划问题.

例 11.2 工厂设址问题

现在准备在 m 个地方建造规模不同的工厂 $A_i(i=1,2,\cdots,m)$，每个工厂的建厂费用为 $f_i(i=1,2,\cdots,m)$，并假定这些工厂生产相同的产品，各工厂生产的产品数量为 $a_i(i=1,2,\cdots,m)$. 工厂附近有 n 个零售商店 $B_j(j=1,2,\cdots,n)$ 承销工厂的产品，每个零售商店对产品的需求量为 $b_j(j=1,2,\cdots,n)$. 设 c_{ij} 表示从 A_i 运送一件产品到 B_j 所需运费，问如何安排建厂计划，也就是建造哪些工厂，能够既满足零售需求，又使得建造工厂和运输产品的总费用最小?

解 此问题需设立两个决策变量：x_{ij} 表示 A_i 运至 B_j 的产品数量，且 $x_{ij} \geqslant 0$ $(i=1,2,\cdots,m; j=1,2,\cdots,n)$；$y_i$ 为 $0-1$ 变量，当在 A_i 处建厂时取 1，当不在 A_i 处建厂时取 0 $(i=1,2,\cdots,m)$.

目标函数为总费用

$$z = \sum_{i=1}^{m} \left[f_i y_i + \sum_{j=1}^{n} c_{ij}x_{ij} \right],$$

该问题的数学模型为

$$\min z = \sum_{i=1}^{m} \left[f_i y_i + \sum_{j=1}^{n} c_{ij}x_{ij} \right],$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{m} x_{ij} \geqslant b_j, j=1,2,\cdots,n, \\ \sum_{j=1}^{n} x_{ij} \leqslant a_i y_i, i=1,2,\cdots,m, \\ y_i = 0 \text{ 或 } 1, \\ x_{ij} \geqslant 0, \text{非负实数} (i=1,2,\cdots,m; j=1,2,\cdots,n). \end{cases}$$

这是一个混合整数规划.

例 11.3 背包问题

设有一个旅行者需要在背包里装入 n 件不同的物品，每件物品的质量分别为 $a_j(j=1,2,\cdots,m)$，价值分别为 $c_j(j=1,2,\cdots,m)$，现在要使背包的总质量不超过 b，则应装入哪些物品使得总价值最大？

解　设决策变量

$$x_j=\begin{cases}1, & \text{当将第 } j \text{ 件物品装入背包时},\\ 0, & \text{当将第 } j \text{ 件物品不装入背包时}.\end{cases}$$

目标函数为总价值 $z=\sum_{j=1}^{n}c_jx_j$.

该问题的数学模型为

$$\max z=\sum_{j=1}^{n}c_jx_j,$$

$$\text{s.t.}\begin{cases}\sum_{j=1}^{n}a_jx_j\leqslant b,\\ x_j=0 \text{ 或者 } 1, j=1,2,\cdots,n.\end{cases}$$

这是一个 0－1 规划问题.

11.2　整数规划的分枝定界法和割平面法

整数规划模型的一般形式为

$$\max\ (\min)\ z=\sum_{j=1}^{n}c_jx_j,$$

$$\text{s.t.}\begin{cases}\sum_{j=1}^{n}a_{ij}x_j\leqslant(\geqslant,=)b_i(i=1,2,\cdots,m),\\ x_j=0 \text{ 或者 } 1\ (j=1,2,\cdots,n).\end{cases}\tag{11-1}$$

对于实际中的某些整数规划问题，很容易会想到先略去整数约束的条件，即视为一个线性规划问题，利用单纯形法求解，然后对其最优解进行取整处理. 实际上，这样得到的解未必是原整数规划问题的最优解，因此，这种方法是不可取的，但可借鉴这种思想.

下面介绍求解整数规划问题的两种常用方法：分枝定界法和割平面法.

11.2.1　分枝定界法

分枝定界法可以解决纯整数或混合的整数规划问题. 分枝就是在相应的无整数约束的线性规划问题的可行域中，任取一个非整数解的非整数变量，对此变量附加两个相互排斥的不等式约束，将原问题分为两个于

问题(即为分枝),再分别求解子问题的解,直到子问题的最优解为原问题的可行解(即整数解)为止.定界就是在分枝过程中,依据每次分枝所得子问题的最优解来确定原问题最优解的上下界,求解过程是该上下界所确定的范围内分枝并逐渐缩小上下界的范围,直到求出原问题的最优解为止.

设要求解的原整数规划问题记为问题(A),而问题(A)去掉整数约束条件后的线性规划问题记为(B),则问题(B)的解与原问题(A)的解有如下的关系:

(1) 若问题(B)没有可行解,则问题(A)也没有可行解;

(2) 若问题(B)有最优解,并符合问题(A)的整数条件,则问题(B)的最优解也是问题(A)的最优解;

(3) 若问题(B)有最优解,但不符合问题(A)的整数条件,则不是问题(A)的最优解.

下面给出分枝定界法的流程图(图 11－1)和算法.

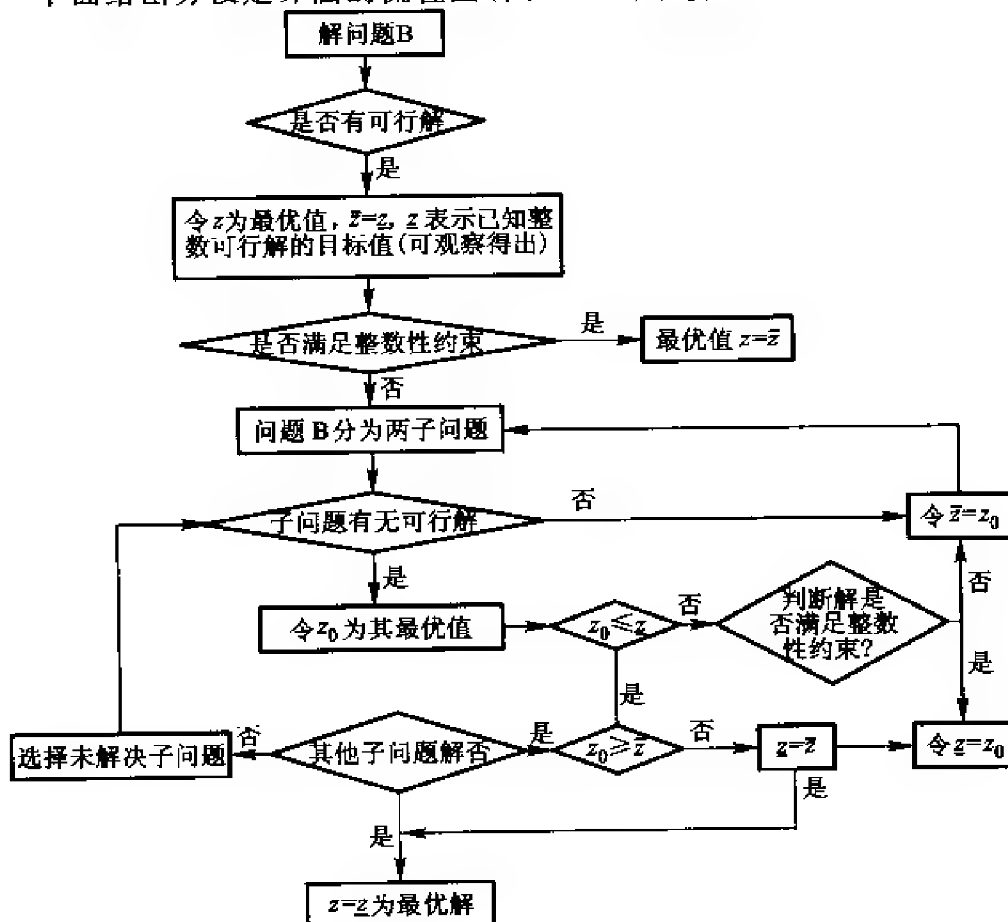


图 11－1　分枝定界法的流程图

算法 11.1

（一）初始步

求解问题（A）去掉整数约束条件后的问题（B）.

① 判断问题（B）的解，若问题（B）没有可行解，这时问题（A）也没有可行解，则停止；若问题（B）有最优解，则转 2°；

② 检查是否符合问题（A）的整数条件？

是，问题（B）的最优解即为问题（A）的最优解，则停止；

否，转③；

（二）迭代步

③ 记问题（B）的目标函数值为问题（A）的最优目标函数 z^* 的上界，记为 $\bar{z}$. 再通过观察法找问题（A）的一个整数可行解，不妨取 $x_j=0\ (j=1,2,\cdots,n)$，求得其目标函数值记作 $\underline{z}$. 这时显然 $\underline{z}\leqslant z^*\leqslant\bar{z}$.

④ 分枝：在问题（B）的最优解中任选一个不符合整数条件的变量 x_j，其值为 b_j，以 $[b_j]$ 表示不超过 b_j 的最大整数，构造两个相互排斥的不等式约束条件 $x_j\leqslant[b_j]$ 和 $x_j\geqslant[b_j]+1$，将这两个约束条件分别加入问题（B）中去，得到两个子问题（B_1）和（B_2）. 用单纯形等方法求出子问题（B_1）和（B_2）的最优解. 再从求得的结果中找出最优值的最大者为新的上界 $\bar{z}$，从已符合整数条件的各分枝中，找出目标函数值为最大者作为新的下界 $\underline{z}$.

⑤ 比较和剪枝：找出各分支（子问题）的最优值有小于下界值 $\underline{z}$ 者，则可以剪掉这个分枝（即去相应的子问题），不再考虑. 若大于 $\underline{z}$，且不符合整数条件，则重复进行第一步骤，直到得到 $z^*=\underline{z}$ 为止.

例 11.4　用分枝定界法求解整数规划问题

$$\max z = 40x_1 + 90x_2,$$

$$\text{s.t.}\begin{cases}9x_1 + 7x_2 \leqslant 56,\\ 7x_1 + 20x_2 \leqslant 70,\\ x_1, x_2 \geqslant 0,\\ x_1, x_2 \text{ 为整数}.\end{cases}$$

解　由原问题（记为（A））去掉整数条件得问题（B）：

$$\max z = 40x_1 + 90x_2,$$

$$\text{s.t.}\begin{cases}9x_1 + 7x_2 \leqslant 56,\\ 7x_1 + 20x_2 \leqslant 70,\\ x_1, x_2 \geqslant 0.\end{cases}$$

求解问题（B）的过程如图 11－2 所示：

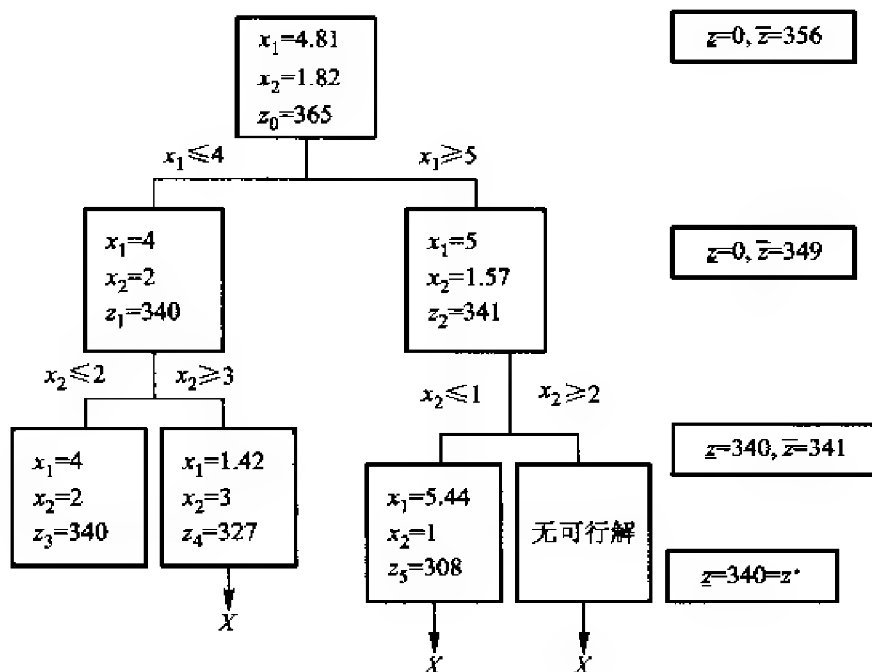


图 11－2

故求得最优解为 $\boldsymbol{x}=(4,2)^{\mathrm{T}}$，最优值为 $z=340$.

11.2.2　割平面法

割平面法是由 R. E. Gomory 在 1958 年首先提出来的，故也称为 Gomory 割平面法. 首先不考虑变量 x_i 为整数的约束条件，适当的增加约束条件（称为 Gomory 条件，几何术语为割平面）使将问题的原可行域切割掉一部分，这部分只包含非整数解，但不包含任何整数可行解，这样可以得到一个新的子问题，然后求解这个子问题的解，直到子问题的最优解为原问题的最优解为止. 下面先用具体的问题来说明割平面法.

例 11.5　用割平面法求解整数规划问题

$$\max z = x_1 + x_2,$$

$$\text{s.t.}\begin{cases}-x_1+x_2\leqslant 1,\\ 3x_1+x_2\leqslant 4,\\ x_1,x_2\geqslant 0,\\ x_1,x_2\text{ 为整数.}\end{cases}$$

解　由原问题（记为(A)）去掉整数条件得问题(B)：

$$\max z = x_1 + x_2,$$

$$\text{s.t.}\begin{cases}-x_1+x_2\leqslant 1,\\3x_1+x_2\leqslant 4,\\x_1,x_2\geqslant 0.\end{cases}$$

对问题(B)由单纯形法可得非整数最优解 $x_1=\frac{3}{4}$,$x_2=\frac{7}{4}$,$z=\frac{5}{2}$,最终单纯形表如表 11－2 所示.

表 11－2

	c_j		1	1	0	0
c_B	x_B	b	x_1	x_2	x_3	x_4
1	x_1	$\frac{3}{4}$	1	0	$-\frac{1}{4}$	$\frac{1}{4}$
1	x_2	$\frac{7}{4}$	0	1	$\frac{3}{4}$	$\frac{1}{4}$
	σ_j		0	0	$-\frac{1}{2}$	$-\frac{1}{2}$

显然得到的最优解不是整数解. 故想找如图 11－3 中的线段 CD 去切割可行域 R,问题就是如何去找"割平面"的线段 CD(如图 11－3).

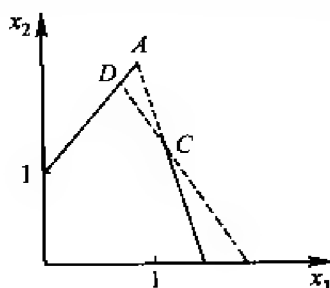


图 11－3

由表 11－2 可得关系式

$$\begin{cases}x_1-\frac{1}{4}x_3+\frac{1}{4}x_4=\frac{3}{4},\\x_2+\frac{3}{4}x_3+\frac{1}{4}x_4=\frac{7}{4}.\end{cases}$$

将系数和常数项都分解成整数和非负真分数之和,则

$$\begin{cases}x_1-x_3+\frac{3}{4}x_3+\frac{1}{4}x_4=\frac{3}{4},\\x_2+\frac{3}{4}x_3+\frac{1}{4}x_4=1+\frac{3}{4},\end{cases}$$

即

$$\begin{cases} x_1 - x_3 = \frac{3}{4} - \left(\frac{3}{4}x_3 + \frac{1}{4}x_4\right), \\ x_2 - 1 = \frac{3}{4} - \left(\frac{3}{4}x_3 + \frac{1}{4}x_4\right). \end{cases} \quad (11-2)$$

现考虑条件(11 -2),要使 x_1,x_2 为非负整数.由

$$\begin{cases} -x_1 + x_2 + x_3 = 1, \\ 3x_1 + x_2 + x_4 = 4, \end{cases}$$

可知 x_3,x_4 也为非负整数.在(11 -2)式中,从等式左边看是整数,等式右边括号内是正数,则等式右边必是负整数,即

$$\frac{3}{4} - \left(\frac{3}{4}x_3 + \frac{1}{4}x_4\right) \leqslant 0,$$

故

$$-3x_3 - x_4 \leqslant -3.$$

这就是一个切割方程,将它作为约束条件增加到问题中去得到一个新的子问题,再进行求解.

引入松弛变量 x_5,得到等式 $-3x_3 - x_4 + x_5 = -3$,将新的约束方程加到最终计算表中.由于 b 列有负数,故得到的是非可行解,于是需要用对偶单纯形法求解(具体见表 11 -3).

表 11 -3

c_j			1	1	0	0	0
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5
1	x_1	$\frac{3}{4}$	1	0	$-\frac{1}{4}$	$\frac{1}{4}$	0
1	x_2	$\frac{7}{4}$	0	1	$\frac{3}{4}$	$\frac{1}{4}$	0
0	x_5	-3	0	0	-3	-1	1
σ_j			0	0	$-\frac{1}{2}$	$-\frac{1}{2}$	0
1	x_1	1	1	0	0	$\frac{1}{3}$	$\frac{1}{12}$

续表

c_j			1	1	0	0	0
c_B	x_B	b	x_1	x_2	x_3	x_4	x_5
1	x_2	1	0	1	0	-1	$\frac{1}{4}$
0	x_3	1	0	0	1	$-\frac{1}{3}$	$-\frac{1}{3}$
σ_j			0	0	0	$-\frac{1}{3}$	$-\frac{1}{6}$

这时得到最优整数解 $x=(1,1)^{\mathrm{T}}$，最优值为 $z=\frac{5}{2}$.

11.3　0－1 规划

11.3.1　0－1 规划举例

在整数规划中有一类特殊的情形，即决策变量只取 0 或 1 两个值，则此规划称为 0－1 规划. 如在 11.1 节中的背包问题(例 11.3)，它的所有决策变量 x_j 只能取值 0 或 1，这时 x_j 称为 0－1 变量. 下面来看 0－1 规划的几个实际问题.

例 11.6　投资问题

现有总额为 B 的资金可用于投资，共有 n 个项目可供投资者选择. 已知项目 j 所需投资额为 a_j，投资后可得利润 $c_j(j=1,2,\cdots,n)$，不妨设 B、a_j、c_j 均是整数. 问如何选择投资项目可使利润最大?

解　先引人 0－1 变量 x_j，令

$$x_j=\begin{cases}1, & \text{对项目 } j \text{ 投资},\\ 0, & \text{否则},\end{cases}$$

便可得如下 0－1 整数规划问题

$$\max z=\sum_{j=1}^{n}c_jx_j,$$

$$\text{s.t.}\begin{cases}\sum_{j=1}^{n}a_jx_j\leqslant B,\\ x_j=0 \text{ 或 } 1, j=1,2,\cdots,n.\end{cases}$$

例 11.7　相互排斥的约束条件

如果有 m 个互相排斥的约束条件(≤型):

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i, i = 1,2,\cdots,m,$$

要求其中至少有 q 组得到满足,而其他 $m-q$ 组约束可以满足,也可以不满足.

解 引入 m 个 0-1 变量 $y_i(i=1,2,\cdots,m)$ 和一个充分大的常数 M,而下面这一组 $m+1$ 个约束条件:

$$\begin{cases} a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i + y_iM, i = 1,2,\cdots,m, \\ y_1 + y_2 + \cdots + y_m \leq m - q \end{cases}$$

就能满足要求,为了使约束条件中至少 q 组成立,则至少有 q 个 y_i 为零.

11.3.2 0-1 规划的解法

1. 完全枚举法

检查变量取值为 0 或 1 的每一组合,比较目标函数值的大小以求得最优解.这种方法虽思路自然,但当 n 比较大时,n 个变量取值的组合数 2^n 相当大,一般不可行.下面以最大化问题为例给出完全枚举法的算法.

算法 11.2

① 按价值系数 c_j 的不减顺序排列目标函数 $z=f(x)$ 和约束条件 ($j=1,2,\cdots,n$);

② 列出所有的解 $x^{(j)}(j=1,2,\cdots,2^n)$,置 $\underline{z}=-M$;

③ 置 $i=1$;

④ 检查 $f(x^{(i)})>\underline{z}$?

是,转⑤,

否,转⑦;

⑤ 检查 $x^{(i)}$ 可行?

是,转⑥,

否,转⑦;

⑥ 置 $f(x^{(i)})\Rightarrow\underline{z}$;

⑦ 置 $i=i+1$;

⑧ 检查 $i=2^n+1$?

是,转⑨,

否,转④;

⑨ 检查 $\underline{z}=-M$?

是,无解,

否,$\max z=\underline{z}$,停止计算.

2. 隐枚举法

这种方法与完全枚举不同的是只检查变量取值组合的一部分，就能求得问题的最优解. 基本思路是：找到一组可行解，然后改进目标值，直到不能改进为止. 其实分枝定界法也是一种隐枚举法，但对有些问题隐枚举法不一定适用.

例 11.8　求解下列整数规划问题

$$\max z = 3x_1 - 2x_2 + 5x_3,$$

$$\text{s.t.}\begin{cases} x_1 + 2x_2 - x_3 \leqslant 2, & (11-3)\\ x_1 + 4x_2 + x_3 \leqslant 4, & (11-4)\\ x_1 + x_2 \leqslant 3, & (11-5)\\ 4x_2 + x_3 \leqslant 6, & (11-6)\\ x_1, x_2, x_3 = 0 \text{ 或 } 1. & (11-7)\end{cases}$$

解　首先通过试探找一个可行解，如 $(x_1, x_2, x_3)^T = (1,0,0)^T$ 满足 (11－3)—(11－6)，得 $z=3$，于是增加一个约束条件

$$3x_1 - 2x_2 + 5x_3 \geqslant 3, \qquad (11-8)$$

称(11－8)为过滤条件. 现将 5 个约束条件按(11－8)，(11－3)—(11－6)的顺序排好，对每个解依次代入约束条件左侧，求出值，如某一条件不满足，同行的以下各条件则不必再检查. 计算过程如表 11－4.

表 11－4

点	条件					满足条件否	z 值
	(11－8)	(11－3)	(11－4)	(11－5)	(11－6)		
(0　0　0)	0					×	
(0　0　1)	5	－1	1	0	1	√	5
(0　1　0)	－2					×	
(0　1　1)	3	1	5			×	
(1　0　0)	3	1	1	1	0	√	3
(1　0　1)	8	0	2	1	1	√	8
(1　1　0)	1					×	
(1　1　1)	6	2	6			×	

所以最优解 $x=(1,0,1)^T$，最优值为 $z=\frac{5}{2}$.

在计算过程中，还可进一步减少计算量，具体有这样一些方法可以更早发现最优解：

（1）若遇到 z 值已超过条件（11－8）右边的值，再改变条件（11－8），如当检查（0,0,1）时，因 $z=5(>3)$，故可把条件（11－8）换成

$$3x_1-2x_2+5x_3\geq 5. \qquad (11-8')$$

（2）重新排列 x_j 的顺序使目标函数中 x_j 的系数是递增的．如此例中，改写 $z=3x_1-2x_2+5x_3=-2x_2+3x_1+5x_3$，因 $-2,3,5$ 是递增的，变量的组合也以 (x_2,x_1,x_3) 的顺序排列，按各变量的递增顺序列出的解有：(0,0,0)，(0,0,1)，(0,1,0)，(0,1,1)，…，按这样的顺序寻优，可以更早发现最优解，再结合过滤条件的改进，使计算更简化．具体计算过程见表 11－5、表 11－6、表 11－7.

表 11－5

点	条件					满足条件否	z 值
	(11－8)	(11－3)	(11－4)	(11－5)	(11－6)		
(0　0　0)	0					×	
(0　0　1)	5	－1	1	0	1	√	5

改进过滤条件，用 $-2x_2+3x_1+5x_3\geq 5$ 代替 $-2x_2+3x_1+5x_3\geq 3$，继续进行

表 11－6

点	条件					满足条件否	z 值
	(11－8)	(11－3)	(11－4)	(11－5)	(11－6)		
(0　1　0)	3					×	
(0　1　1)	8	0	2	1	1	√	8

再改进过滤条件，用 $-2x_2+3x_1+5x_3\geq 8$ 代替 $-2x_2+3x_1+5x_3\geq 5$，再继续

表 11－7

点	条件					满足条件否	z 值
	(11－8)	(11－3)	(11－4)	(11－5)	(11－6)		
(1　0　0)	2					×	
(1　0　1)	3					×	
(1　1　0)	1					×	
(1　1　1)	6					×	

（3）对于 n 个变量的一般问题，有 2^n 组不同取值，此方法仍无法实际应用，因无法检查全部 2^n 个点. 这时，可从目标值最大的点试起，只要发现使目标最大的可行解，即停止搜索. 为此，需变换目标函数，使各项系数均非负. 如上例中，x_2 系数 -2，用 $(1-\bar{x}_2)$ 代替 x_2，改写成 $3x_1+2\bar{x}_2+5x_3-2$，搜索从所有 x_i 和 $\bar{x}_i$ 取 1 开始（min 问题从取 0 开始）. 约束条件中 x_2 亦用 $\bar{x}_2$ 代替，搜索时先从检验系数大的变量取 1 的点，在上例中，先检查 $(1,1,1)$，即 $x_1=1,\bar{x}_2=1,x_3=1$，由于此点为可行点，终止搜索. 原问题最优解为 $\boldsymbol{x}=(1,0,1)^{\mathrm{T}}$，目标函数最优值为 $z=8$. 这种算法能否真正求得最优解仍无法保证，全靠碰运气，但在求解实例时如能获得成功，至少可求得最优解. 如果一个隐枚举法能对绝大多数实例在较快时间内求出最优解，它仍不失为一个好算法.

11.4 指（分）派问题

11.4.1 指（分）派问题举例

例 11.9 设有 n 个工人 $\mathrm{A}_i(i=1,2,\cdots,n)$，现在要指派他们分别完成 n 项不同的工作 $\mathrm{B}_j(j=1,2,\cdots,n)$，设工人 A_i 完成工作 B_i 所消耗的时间为 $c_{ij}(i,j=1,2,\cdots,n)$（如表 11－8）. 问如何制定这分配方案使完成全部工作的总工时最少？

表 11－8

	B_1	B_2	$\cdots$	B_n
A_1	c_{11}	c_{12}	$\cdots$	c_{1n}
A_2	c_{21}	c_{22}	$\cdots$	c_{2n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
A_n	c_{n1}	c_{n2}	$\cdots$	c_{nn}

解 设决策变量 x_{ij} 表示 B_j 是否分给 A_i 的状况，即每人对每件工作只有做与不做两种情况：

$$x_{ij}=\begin{cases}1, & \mathrm{B}_j \text{ 分给 } \mathrm{A}_i, \\ 0, & \mathrm{B}_j \text{ 不分给 } \mathrm{A}_i.\end{cases}$$

目标函数为总工时：$z=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}c_{ij}x_{ij}$，

约束条件：$\sum\limits_{i=1}^{n}x_{ij}=1, j=1,\cdots,n$，每件工作只能分派一人做；

$$\sum_{j=1}^{n} x_{ij} = 1, i = 1, \cdots, n, \text{每人只做一件工作.}$$

该问题的数学模型为

$$\min z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij},$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{n} x_{ij} = 1, j = 1, \cdots, n, \\ \sum_{j=1}^{n} x_{ij} = 1, i = 1, \cdots, n, \\ x_{ij} = 0 \text{ 或 } 1, i = 1, \cdots, n; j = 1, \cdots, n. \end{cases}$$

上述问题称为指派问题，也称为分派问题，它是一个 0－1 规划，也就是整数规划的特例.

由 c_{ij}组成的系数矩阵$(c_{ij})_{n\times n}$称为效率矩阵，只要效率矩阵$(c_{ij})_{n\times n}$给定了，指派问题也就相应确定了. 由于指派问题中每一可行解的 n^2 个分量中非零分量的个数仅为 n 个，具有高度退化的性质，所以可采用更简便的解法——匈牙利法.

11.4.2 匈牙利法

匈牙利法是 1955 年由 W. W. Kuhn 提出来专门求解指派问题的解法. 由于他引用了匈牙利数学家 D. Konig 一个关于矩阵中零元素的定理：系数矩阵中独立零元素的最多个数等于能覆盖所有零元素的最小直线数，由此把此解法称为匈牙利法.

指派问题的最优解具有如下性质：

性质 若从系数矩阵$(c_{ij})_{n\times n}$的一行（列）各元素中分别减去该行（列）的最小元素，得到新矩阵$(b_{ij})_{n\times n}$，那么以$(b_{ij})_{n\times n}$为系数矩阵求得的最优解和用原系数矩阵求得的最优解相同.

利用这个性质，在最优解保持不变的情况下，原问题的系数矩阵$(c_{ij})_{n\times n}$可以变换为有很多零元素的新系数矩阵$(b_{ij})_{n\times n}$，在新系数矩阵$(b_{ij})_{n\times n}$中，找 n 个独立的零元素，即找位于不同行不同列的零元素，令解矩阵$(x_{ij})_{n\times n}$中对应的这 n 个独立的零元素的元素取值为 1，其他元素取值为零，将其代入目标函数得到 $z=0$，这显然一定是以$(b_{ij})_{n\times n}$为系数矩阵的指派问题的最优解，也就得到了原问题的最优解.

下面是匈牙利法的算法：

算法 11.3

（一）初始步

变换指派问题的系数矩阵$(c_{ij})_{n\times n}$,使变换后的矩阵$(b_{ij})_{n\times n}$中的各行各列中出现零元素.即对$(c_{ij})_{n\times n}$的每行每列的各元素分别减去该行该列的最小元素,若每行每列已有零元素,则不必再减.

(二) 迭代步

① 进行试指派,寻求最优方案.

(1) 从只有一个零元素的行(列)开始,对这个零元素标记⓪,表示对这行所代表的人只有一种任务可指派,再划去⓪所在的列(行)的其他零元素,标记 Ø,表示此列所代表的任务已经指派完,不考虑别人;

(2) 从只有一个零元素的列(行)开始,对这个零元素标记⓪,再划去加圈零元素所在的行(列)的其他零元素,标记 Ø,反复进行,直到所有零元素都已圈出或划掉.

(3) 若仍有没有划圈的零元素,且同行(列)的零元素至少有两个,则从零元素最少的行(列)开始,通常方法是比较这行各零元素所在列中零元素的数目,选零元素少的那列的这个零元素加圈,然后划掉同行同列的其他元素.反复进行,直到所有的零元素都已圈出或划掉.

如果圈的元素的数目 m = 矩阵阶数 n,则得到最优解,停止;

如果圈的元素的数目 m < 矩阵阶数 n,则转入②;

② 作最少的直线覆盖所有的零元素,以确定该系数矩阵中能找到最多的独立元素数,具体方法为:

(1) 对没有⓪的行打 ✓ 号;

(2) 对已经打 ✓ 号的行中所有含零元素的列打 ✓ 号;

(3) 再对打 ✓ 号的列中含⓪元素的行打 ✓ 号,重复以上步骤直到没有新的打 ✓ 号的行列为止;

(4) 对没有打 ✓ 号的行画横线,对所有打 ✓ 号的列画竖线.这得到覆盖所有零元素的最少直线数目.

如果直线的数目 l = 矩阵阶数 n,则转入②的第(3)步,重新试指派;

如果直线的数目 l < 矩阵阶数 n,则转入③;

③ 为找到 n 个独立的零元素,必须再变换当前系数矩阵以增加零元素.为此在没有被直线覆盖的部分中找出最小元素,在打 ✓ 号的各行中元素都减去这最小元素,在打 V 号的各列的元素都加上这最小元素,以保证原来零元素不变,得到新系数矩阵,它的最优解也和原问题相同.

如果得到 n 个独立的零元素,则得到最优解,停止;

否则,转入②,重复进行.

例 11.10　有 4 个工人,要指派他们分别完成 4 项工作,每人做各项

工作所消耗的时间如下表. 问指派哪个人去完成哪项工作,可使总的消耗时间最小?

人员 \ 任务	A	B	C	D
甲	15	18	21	24
乙	19	23	22	18
丙	26	17	16	19
丁	18	21	23	17

解

$$(c_{ij})_{n\times n}=\begin{bmatrix}15&18&21&24\\19&23&22&18\\26&17&16&19\\19&21&23&17\end{bmatrix}\begin{matrix}-15\\-18\\-16\\-17\end{matrix}\to\begin{bmatrix}0&3&6&9\\1&5&4&0\\10&1&0&3\\2&4&6&0\end{bmatrix}$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad -1$$

$$\to\begin{bmatrix}0&2&6&9\\1&4&4&0\\10&0&0&3\\2&3&6&0\end{bmatrix}=(b_{ij})_{n\times n}.$$

$$(b_{ij})_{n\times n}=\begin{bmatrix}⓪&2&6&9\\1&4&4&⓪\\10&⓪&\emptyset&3\\2&3&6&\emptyset\end{bmatrix}$$

这里⓪的个数 $m=3,n=4$,

$$\begin{bmatrix}⓪&2&6&9\\1&4&4&⓪\\10&⓪&\emptyset&3\\2&3&6&\emptyset\end{bmatrix}\begin{matrix}\\ \vee\\ \\ \vee\end{matrix},$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\vee$$

故得到直线数目 $l=3<n$,在没有被覆盖部分(第 2,4 行)中找出最小元素 1,然后在第 2,4 行各元素分别减少 1,给第 4 列各元素加 1,得到新矩

阵，重复上面步骤，得

$$\begin{bmatrix} \textcircled{0} & 2 & 6 & 9 \\ \not{0} & 3 & 3 & \textcircled{0} \\ 10 & \textcircled{0} & \not{0} & 3 \\ 2 & 3 & 6 & \not{0} \end{bmatrix}\begin{matrix} \\ \checkmark \\ \\ \checkmark \end{matrix},$$
$$\begin{matrix} \checkmark & & & \checkmark \end{matrix}$$

这里(0)的个数 $m=3,n=4$，

$$\begin{bmatrix} \not{0} & 2 & 6 & 10 \\ \not{0} & 3 & 3 & \textcircled{0} \\ 10 & \not{0} & \textcircled{0} & 4 \\ \textcircled{0} & 2 & 5 & \not{0} \end{bmatrix}\begin{matrix} \checkmark \\ \checkmark \\ \\ \checkmark \end{matrix},$$
$$\begin{matrix} \checkmark & & & \checkmark \end{matrix}$$

故得到直线数目 $l=3<n$，在没有被覆盖部分中找出最小元素 2，在第 1，2，4 行中各元素分别减去 2，给第 1，4 列各元素加 2，得新矩阵

$$\begin{bmatrix} \not{0} & \textcircled{0} & 4 & 12 \\ \not{0} & 1 & 1 & \textcircled{0} \\ 12 & \not{0} & \textcircled{0} & 6 \\ \textcircled{0} & \not{0} & 3 & \not{0} \end{bmatrix},$$

它具有 4 个独立的 0 元素，这就得到最优解，相应的矩阵为

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

由解矩阵得到最优指派方案：

甲 ——B，乙 ——D，丙 ——C，丁 ——A，

所需时间为 $\min z=18+18+16+18=70$.

11.5　整数规划问题应用实例

例 11.11　石材截断切割问题（97 年全国大学生数学建模竞赛题）

设将一长方体石材，加工成另一长方体石材．石材母体：长 $L=10$ cm，宽 $w=14.5$ cm，高 $h=19$ cm；石材子体：长 $L=3$ cm，宽 $w=2$ cm，高 $h=4$ cm；加工要求如下：

（1）母体与子体的对应表面平行；

（2）间距分别为：左侧间距 6 cm，右侧间距 1 cm，正面间距 7 cm，后面间距 5.5 cm，底面间距 9 cm，上面间距 6 cm；假定只要求切割 6 次，加工费用 1 元/厘米^2，试设计一种方案（即安排各面加工次序），使加工费用最少？

解　首先给出各种不同切割方式的数学表述.

设 6 个加工面的编号分别为：

1	2	3	4	5	6
左	右	前	后	上	下

一个切割方式就是各加工面{1,2,3,4,5,6}的一个全排列，记为

$$i_1,i_2,i_3,i_4,i_5,i_6,$$

用 R 记所有不同的切割方式组成的集合，即

$$R=\{(i_1,i_2,i_3,i_4,i_5,i_6)\mid i_j=1,2,3,4,5,6,$$

$$j=1,2,3,4,5,6,i_j\neq i_k(k\neq j)\}=6!=720\text{ 种},$$

然后列出 6 次切割的总费用的数学描述：

现考虑任一种切割方式 i_1,i_2,i_3,i_4,i_5,i_6 的费用描述，由于这种切割方式在操作中总共要进行 6 次切割，记 6 次切割的面积分别为(cm^2)：

$$S(i_1),S(i_1,i_2),S(i_1,i_2,i_3),S(i_1,i_2,i_3,i_4),$$

$$S(i_1,i_2,i_3,i_4,i_5),S(i_1,i_2,i_3,i_4,i_5,i_6),$$

故 6 次切割的总费用为

$$f(i_1,i_2,i_3,i_4,i_5,i_6)=\sum_{j=1}^{6}S(i_1,i_2,\cdots,i_j).$$

最后列出该问题的数学模型

$$\min z=f(i_1,i_2,i_3,i_4,i_5,i_6)$$

$$\text{s.t.}\quad i_1,i_2,i_3,i_4,i_5,i_6\in R$$

这是一个典型的整数规划问题，我们给出用穷举法求解该问题的算法步骤：

① 列出所有可能的切割方案(720 种)；

② 计算每一种切割方案的总费用(720 个)；

③ 从 720 个总费用中选出最小者；

④ 列出最少费用方案(最优方案)；

计算结果如下：

按如下两种方案(各面加工次序)加工，最少费用均为 374 元：

第一种：1 底(6)，前(3)，左(1)，上(5)，后(4)，右(2)；

第二种：2 底(6)，前(3)，上(5)，左(1)，后(4)，后(2).

第 12 章　约束最优化问题的最优性条件

12.1　约束最优化问题的数学描述

不失一般性，设我们讨论如下的约束极小问题（(PP)问题）：

$$\begin{aligned} \min \quad & f(\boldsymbol{x}),\boldsymbol{x} \in \mathbf{R}^n, \\ \text{s.t.} \quad & g(\boldsymbol{x}) \geqslant 0, \\ & h(\boldsymbol{x}) = 0, \end{aligned} \tag{12-1}$$

其中 $f(\boldsymbol{x}),g(\boldsymbol{x})=(g_1(\boldsymbol{x}),g_2(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x}))^{\mathrm{T}},h(\boldsymbol{x})=(h_1(\boldsymbol{x}),h_2(\boldsymbol{x}),\cdots,h_l(\boldsymbol{x}))^{\mathrm{T}}$，$f(\boldsymbol{x}),g_i(\boldsymbol{x})(i=1,2,\cdots,m),h_i(\boldsymbol{x})(i=1,2,\cdots,l)$ 均为定义在 $X\subset\mathbf{R}^n$ 上的实值函数.

显然，$l=m=0$ 时，(12－1) 便是一个无约束最优化问题，当 $f(\boldsymbol{x})$，$g(\boldsymbol{x}),h(\boldsymbol{x})$ 均为线性函数时，它是一个线性规划问题. 这两种情形前面已经介绍过了，从本章起我们假定 l,m 中至少有一个不为 0，且 $f(\boldsymbol{x}),g(\boldsymbol{x}),h(\boldsymbol{x})$ 中至少有一个为非线性函数. 此时称问题(12－1)为非线性规划.

12.1.1　全局解与局部解

定义 12.1　称集合

$D=\{\boldsymbol{x}\mid h_i(\boldsymbol{x})=0,i=1,2,\cdots,l,g_i(\boldsymbol{x})\geqslant 0,i=1,2,\cdots,m,\boldsymbol{x}\in\mathbf{R}^n\}$

为问题(12－1)的可行集（或称可行域、容许集）.

定义 12.2　对于问题(12－1)若存在 $\boldsymbol{x}^*\in D$，使得对任意的 $\boldsymbol{x}\in D$，均有 $f(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*)$，则称 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 在可行域上的全局极小点，或称 $\boldsymbol{x}^*$ 是问题(12－1)的全局最优解. 若对任意的 $\boldsymbol{x}\in D,\boldsymbol{x}\neq\boldsymbol{x}^*$，均有 $f(\boldsymbol{x})>f(\boldsymbol{x}^*)$，则称 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 在可行域上的严格全局极小点，或称 $\boldsymbol{x}^*$ 是问题(12－1)的严格全局最优解.

定义 12.3　对于问题(12－1)，若对 $\boldsymbol{x}^*\in D$，存在 $\varepsilon>0$，使当 $\|\boldsymbol{x}-\boldsymbol{x}^*\|\leqslant\varepsilon$ 时，总有 $f(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*)$，则称 $\boldsymbol{x}^*$ 为问题(12－1)的局部最优解，或简称 $\boldsymbol{x}^*$ 是最优解. 若当 $\boldsymbol{x}\in D$，且 $0<\|\boldsymbol{x}-\boldsymbol{x}^*\|\leqslant\varepsilon$，总有 $f(\boldsymbol{x})>f(\boldsymbol{x}^*)$，则称 $\boldsymbol{x}^*$ 是问题(12－1)的严格局部最优解.

例 12.1　试确定约束极小问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = x_1^2 + x_2^2, \\ \text{s.t.} \quad & h(\boldsymbol{x}) = (x_1 - 1)^2 - x_2^2 - 4 = 0, \\ & g(\boldsymbol{x}) = x_2 - 1 \leqslant 0 \end{aligned}$$

的局部解.

解　由等式约束与不等式约束的图形(图 12-1)知,可行域即是在 $x_2 \leqslant 1$ 下方且位于双曲线上的两段曲线,而目标函数 $f(\boldsymbol{x}) = x_1^2 + x_2^2$,作它的等值线,发现它们是一系列的同心圆,显然,$\hat{\boldsymbol{x}} = (3,0)^{\mathrm{T}}$ 与 $\boldsymbol{x}^* = (-1,0)^{\mathrm{T}}$ 分别为 $9 = x_1^2 + x_2^2$ 和 $1 = x_1^2 + x_2^2$ 与双曲线的切点,且易知,对任意 $\boldsymbol{x} \in D$,都有 $f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*)$,所以,$\boldsymbol{x}^* = (-1,0)^{\mathrm{T}}$ 为全局极小解,而 $\hat{\boldsymbol{x}} = (3,0)^{\mathrm{T}}$ 为局部极小解.

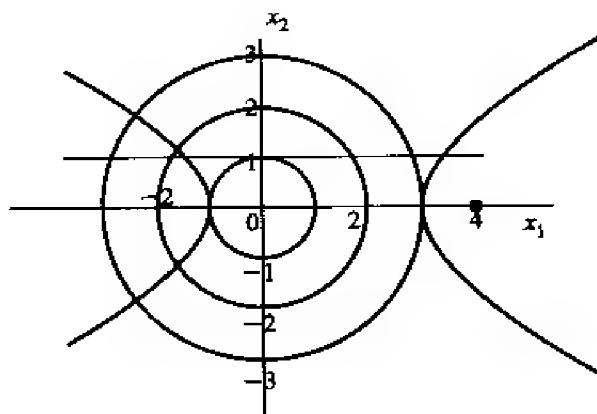


图 12-1

一般说来,对于一个约束极小问题可能出现的情形有:

(1) 若 $D = \varnothing$,则问题无可行解(容许解);

(2) 若 $D \neq \varnothing$,$\inf\limits_{x \in D} \{f(\boldsymbol{x})\}$ 有界,存在 $\boldsymbol{x}^*$ 使 $f(\boldsymbol{x}^*) = \min\limits_{x \in D} \{f(\boldsymbol{x})\} = \inf\limits_{x \in D} \{f(\boldsymbol{x})\}$,则问题有精确解;

(3) 若 $D \neq \varnothing$,$\inf\limits_{x \in D} \{f(\boldsymbol{x})\} = f^0$ 有界,但不存在 $\boldsymbol{x}^0 \in D$ 使 $f(\boldsymbol{x}^0) = f^0$,则问题无精确解;

(4) 若 $D \neq \varnothing$,$\inf\limits_{x \in D} \{f(\boldsymbol{x})\} = -\infty$,则问题为无界解.

12.1.2　凸规划

由于凸规划是一类重要的约束最优化问题,尤其在理论上有其特殊的作用,因此我们有必要引进凸规划的概念并简要介绍其性质.

定义 12.4　对于问题(12-1),若 $f(\boldsymbol{x})$,$h_i(\boldsymbol{x})$ $(i = 1,2,\cdots,l)$ 是开凸

集 S 上的凸(或凹)函数，$g_i(x)(i=1,2,\cdots,m)$ 是线性函数，则问题称为凸规划. 特别的，若 $f(x)$ 还是开凸集 S 上的严格凸函数，则称问题(12－1)是一个严格凸规划问题.

例 12.2　设 $A\in\mathbf{R}^{m\times n}$，$b\in\mathbf{R}^m$，$c\in\mathbf{R}^n$，则线性规划问题

$$\begin{cases}\min & c^{\mathrm{T}}x,\\ \text{s.t.} & Ax=b,\\ & x\geqslant 0\end{cases}$$

是凸规划问题.

解　事实上，线性规划问题的目标函数 $c^{\mathrm{T}}x$ 是线性函数，所以是凸函数，又因为其可行域为

$$D=\{x\in\mathbf{R}^n \mid Ax=b, x\geqslant 0\},$$

它是个多胞形，所以 D 是凸集，从而此线性规划是凸规划.

凸规划较一般非线性规划重要，是因为它具有下面的性质：

定理 12.1　凸规划问题的任何局部极小点都是全局极小点，且极小点的集合为凸集.

证　设 x^* 为一局部极小点，但非全局极小点，则必有可行点 $\bar{x}$ 使 $f(\bar{x})<f(x^*)$. 因为凸规划的可行域是一凸集，所以连线段上的点 $x=(1-\lambda)x^*+\lambda\bar{x}(0<\lambda<1)$ 也都是可行的，又由于 $f(x)$ 的凸性，对于这种点有

$$\begin{aligned}f(x)&\leqslant(1-\lambda)f(x^*)+\lambda f(\bar{x})\\&<(1-\lambda)f(x^*)+\lambda f(x^*)=f(x^*),\end{aligned}$$

所以对于充分小的 $\lambda>0$，也即对于位于 x^* 邻近的连线段上的 x，都有 $f(x)<f(x^*)$，这与 x^* 为局部极小点的假设矛盾，所以凸规划的局部极小点都是全局极小点.

设 $\bar{x}$ 与 $\tilde{x}$ 是两个不相同的全局极小点，自然有 $f(\bar{x})=f(\tilde{x})$，线段

$$x=(1-\lambda)\bar{x}+\lambda\tilde{x}\quad(0<\lambda<1)\tag{12－2}$$

上的点 x 也都是可行的，故由 $f(x)$ 的凸性性质得

$$f(x)\leqslant(1-\lambda)f(\bar{x})+\lambda f(\tilde{x})=f(\bar{x}),$$

又由于 $\bar{x}$ 是全局极小点，所以 $f(x)\geqslant f(\bar{x})$，合起来即得

$$f(x)=f(\bar{x})=f(\tilde{x}),$$

所以，$x\in D$，即 D 是一凸集.

推论 12.1　若凸规划问题为严格凸规划，则其全局极小点（如果存在的话）是唯一的.

证　（反证法）若全局极小点不唯一，则可假设 $\bar{x}$ 与 $\tilde{x}$ 是两个不相同的全局极小点，由定理 12.1，线段（12－2）上的点 x 也都是全局极小点；另一方面，由 $f(x)$ 的严格凸性

可知　$$f(x) < (1-\lambda)f(\bar{x}) + \lambda f(\tilde{x}) = f(\bar{x}) = f(x),$$
从而导出矛盾，结论得证.

12.2　几何最优性条件

考虑非线性约束问题（12－1），我们知道几何最优性条件相对直观，故先给出其问题的几何最优性条件，并辅以图形来理解最优性条件的几何意义，然后再讨论它的代数表示，为此引入几个概念.

以下我们用 $\mathrm{cl}\, D = \{x \in \mathbf{R}^n \mid D \cap N_\delta(x) \neq \varnothing, \forall \delta > 0\}$ 来表示 $\mathbf{R}^n$ 中集合 D 的闭包.

定义 12.5　设 $D \subseteq \mathbf{R}^n$，$\bar{x} \in \mathrm{cl}\, D$，$d \neq 0$，若存在 $\delta > 0$，使得
$$\bar{x} + \lambda d \in D, \forall \lambda \in (0, \delta),$$
则称 d 为集合 D 在点 $\bar{x}$ 处的可行方向.

定义 12.6　如果 d 既是 D 在点 $\bar{x}$ 处的可行方向，又是 $f: D \to \mathbf{R}$ 在点 $\bar{x}$ 处的下降方向，则称 d 为 f 在点 $\bar{x}$ 处的下降可行方向.

由此定义可知，从点 $\bar{x}$ 出发，沿可行方向 d 的微小移动，必能达到 D 的一个点.

定义 12.7　集合 D 在点 $\bar{x}$ 处所有可行方向的集合
$$G = \{d \mid d \neq 0, \exists \delta > 0, \exists \bar{x} + \lambda d \in D, \forall \lambda \in (0, \delta)\}$$
显然是一个以 $\bar{x}$ 为顶点的锥，我们称之为 D 在 $\bar{x}$ 处的可行方向锥.

特别地，当 $\bar{x} \in \mathrm{in}\, D$ 时，D 在点 $\bar{x}$ 处的可行方向锥是全空间 $\mathbf{R}^n$.

定义 12.8　根据第一章定义 1.11，如果 $D \subseteq \mathbf{R}^n$，$\bar{x} \in D$，$d \in \mathbf{R}^n$ 且 $f: D \to \mathbf{R}$ 在点 $\bar{x}$ 处可微，$\nabla f(\bar{x})^{\mathrm{T}} d < 0$，则 d 为 f 在点 $\bar{x}$ 处的下降方向，因此，我们称锥
$$F_0 = \{d \mid \nabla f(\bar{x})^{\mathrm{T}} d < 0\}$$
为 f 在点 $\bar{x}$ 处的下降方向锥.

12.2.1　必要条件

先考虑只有不等式约束的问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}),\boldsymbol{x} \in \mathbf{R}^n, \\ \text{s.t.} \quad & g(\boldsymbol{x}) \geqslant 0, \end{aligned} \tag{12-3}$$

其中 $f(\boldsymbol{x}),g(\boldsymbol{x})=(g_1(\boldsymbol{x}),g_2(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x}))^{\mathrm{T}},f(\boldsymbol{x}),g_i(\boldsymbol{x})(i=1,2,\cdots,m)$ 均为定义在 $X\subset\mathbf{R}^n$ 上的实值函数.

图(12-2)画出了可行方向锥 G 和下降方向锥 F_0,从几何图形上看,若 $\bar{x}$ 不是 D 中极小点,则在点 $\bar{x}$ 处沿下降方向移动,就导致目标函数 f 的值减少;若 $\bar{x}$ 是极小点,任何下降方向都不是可行方向,而任何可行方向也不是下降方向,也即,不存在可行下降方向,这样就得到了一般约束极小问题极小点的一个必要条件.

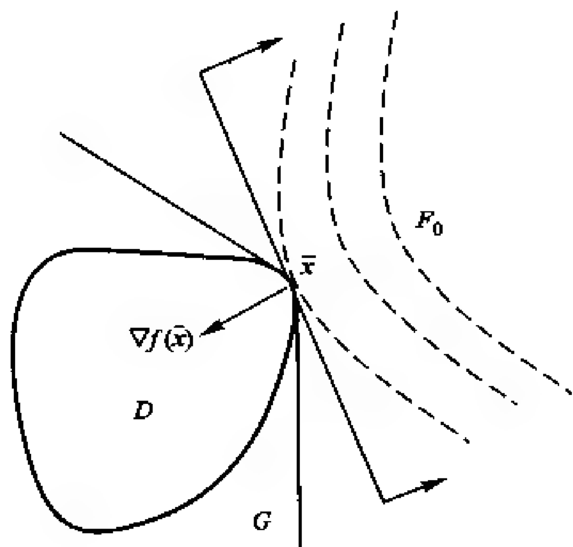


图 12-2

定理 12.2 设约束极小问题

$$\min_{x\in D} f(\boldsymbol{x}), \tag{12-4}$$

其中 $D\subseteq\mathbf{R}^n$ 是非空集合,$f:D\to\mathbf{R}$. 设 $\boldsymbol{x}^*\in D$,且 f 在 $\boldsymbol{x}^*$ 处可微,若 $\boldsymbol{x}^*$ 是该问题的局部极小点,则

$$F_0 \cap G = \varnothing \tag{12-5}$$

证 用反证法. 若存在 $\boldsymbol{d}\in F_0\cap G$,则 $\boldsymbol{d}\in F,\boldsymbol{d}\in G,\boldsymbol{d}\neq 0$. 由 F_0 的定义和第一章定义 1.11,存在 $\delta_1>0$,使

$$f(\boldsymbol{x}^*+\lambda\boldsymbol{d})<f(\boldsymbol{x}^*),\ \forall\lambda\in(0,\delta_1).$$

由 G 的定义,存在 $\delta_2>0$,使

$$\boldsymbol{x}^*+\lambda\boldsymbol{d}\in D,\ \forall\lambda\in(0,\delta_2),$$

再由 $\boldsymbol{x}^*$ 为局部极小点的定义,存在 $\delta_3>0$,使

$$f(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*),\ \forall\boldsymbol{x}\in N_{\delta_3}(\boldsymbol{x}^*)\cap D. \tag{12-6}$$

取 $\delta=\min\{\delta_1,\delta_2,\delta_3\}$，则 $x^*+\frac{\delta}{2}d\in N_{\delta_3}(x^*)\cap D$，且

$$f(x^*+\frac{\delta}{2}d)<f(x^*),$$

此与(12－5)式矛盾.

定义 12.9　设 $\bar{x}$ 为问题(12－3)的可行点，则其不等式约束条件在 $\bar{x}$ 处呈现出两种情形：

(1) $g_i(\bar{x})=0$，称第 i 个不等式约束为在 $\bar{x}$ 处起作用的约束或紧约束；

(2) $g_i(\bar{x})>0$，称第 i 个不等式约束为在 $\bar{x}$ 处不起作用的约束或松约束.

我们用 $I(\bar{x})$ 表示在可行点 $\bar{x}$ 处起作用约束的指标集，即

$$I(\bar{x})=\{i\mid g_i(\bar{x})=0,i=1,2,\cdots,m\}.$$

对于起作用约束，当点沿某些方向稍微离开 $\bar{x}$ 时，仍能满足这些约束，而沿另一些方向离开 $\bar{x}$ 时，不论步长多么小，都会违背这些约束. 对于不起作用约束，当点稍微离开 $\bar{x}$ 时，不论什么方向都不违背这些约束. 如图(12－3)，在 $\bar{x}$ 处，$g_1\geqslant 0$ 和 $g_3\geqslant 0$ 是起作用约束，$g_2\geqslant 0$ 是不起作用约束.

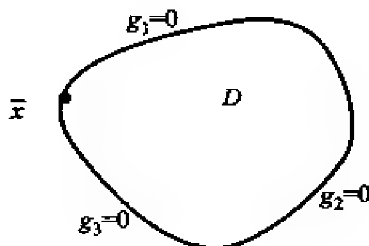


图 12－3

因此，在研究一点处的可行方向时，只需考虑在该点起作用约束，那些不起作用约束可以暂且不管.

借助定理 12.2，下面给出另一个几何最优性条件.

定理 12.3　设 x^* 是问题(12－3)的可行点，f 和 $g_i(i\in I(x^*))$ 在点 x^* 处可微，$g_i(i\notin I(x^*))$ 在点 x^* 处连续，如果 x^* 是该问题的局部极小点，则

$$F_0\cap G_0=\varnothing, \tag{12－7}$$

其中

$$G_0=\{d\mid \nabla g_i(x^*)^{\mathrm{T}}d>0,i\in I(x^*)\}.$$

证　由定理 12.2 知，只需证明 $G_0\subseteq G$.

对于任意 $d \in G_0$,当 $i \in I(x^*)$ 时,$\nabla g_i(x^*)^{\mathrm{T}} d > 0$,所以由下降方向的定义知,存在 $\delta_i > 0$,使得对一切 $\lambda \in (0, \delta_i)$,有

$$g_i(x^* + \lambda d) > g_i(x^*) = 0, i \in I(x^*),$$

而当 $i \notin I(x^*)$ 时,因为 $g_i(x^*) > 0$,所以由 $g_i(x)$ 在点 x^* 处连续知,存在 $\delta_i > 0$,使得对一切 $\lambda \in (0, \delta_i)$,有

$$g_i(x^* + \lambda d) > 0, i \notin I(x^*).$$

取 $\delta = \min\{\delta_i | i = 1, 2, \cdots, m\}$,则对一切 $\lambda \in (0, \delta)$,有

$$g_i(x^* + \lambda d) > 0, i = 1, 2, \cdots, m,$$

从而 $x^* + \lambda d$ 为问题(12-3)的可行点. 显然 $d \neq 0$,因此 d 是在点 x^* 处的一个可行方向,即 $d \in G$,从而 $G_0 \subseteq G$.

对于一般的问题(12-1),有如下定理:

定理 12.4 设 x^* 是问题(12-1)的可行点,f 和 $g_i(i \in I(x^*))$ 在点 x^* 处可微,$g_i(i \notin I(x^*))$ 在点 x^* 处连续,$h_i(x), i = 1, 2, \cdots, l$ 点 x^* 处连续可微,且 $\nabla h_1(x^*), \nabla h_2(x^*), \cdots, \nabla h_l(x^*)$ 线性无关,如果 x^* 是该问题的局部极小点,则

$$F_0 \cap G_0 \cap H_0 = \varnothing, \tag{12-8}$$

其中

$$G_0 = \{d | \nabla g_i(x^*)^{\mathrm{T}} d > 0, i \in I(x^*)\},$$

$$H_0 = \{d | \nabla h_i(x^*)^{\mathrm{T}} d = 0, i = 1, 2, \cdots, l\}.$$

条件(12-8)称为问题(12-1)的几何最优性条件. 易知,此最优性条件仅是必要条件而非充分条件.

例 12.3 设问题为

$$\min\{(x_1 - 4)^2 + (x_2 - 5)^2\},$$

$$\text{s.t.} \begin{cases} 4 - \dfrac{x_1^2}{4} - x_2 \geqslant 0, \\ 8 - x_1 - 2x_2 \geqslant 0, \\ x_1 \geqslant 0, \\ x_2 \geqslant 0, \end{cases}$$

试验证最优点 $x^* = [2, 3]^{\mathrm{T}}$ 满足几何最优性条件,而 $\tilde{x} = [4, 0]^{\mathrm{T}}$ 不满足几何最优性条件.

解 设

$$f(x) = (x_1 - 4)^2 + (x_2 - 5)^2,$$

$$g_1(x) = 4 - \frac{x_1^2}{4} - x_2,$$

$$g_2(\boldsymbol{x}) = 8 - x_1 - 2x_2,$$
$$g_3(\boldsymbol{x}) = x_1,$$
$$g_4(\boldsymbol{x}) = x_2,$$

求梯度

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 4) \\ 2(x_2 - 5) \end{bmatrix}, \nabla g_1(\boldsymbol{x}) = \begin{bmatrix} -\dfrac{x_1}{2} \\ -1 \end{bmatrix}, \nabla g_2(\boldsymbol{x}) = \begin{bmatrix} -1 \\ -2 \end{bmatrix},$$

$$\nabla g_3(\boldsymbol{x}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \nabla g_4(\boldsymbol{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

在点 $\boldsymbol{x}^* = [2,3]^T$ 处,因为 $g_1(\boldsymbol{x}^*) = 0$ 和 $g_2(\boldsymbol{x}^*) = 0$,所以 $g_1(\boldsymbol{x}) \geqslant 0$ 和 $g_2(\boldsymbol{x}) \geqslant 0$ 是起作用约束;$I = \{1,2\}$. 我们来说明,不存在同时满足以下三个不等式的向量:

$$\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} = [-4, -4]\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -4d_1 - 4d_2 < 0,$$

$$\nabla g_1(\boldsymbol{x}^*)^T \boldsymbol{d} = [-1, -1]\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -d_1 - d_2 > 0,$$

$$\nabla g_2(\boldsymbol{x}^*)^T \boldsymbol{d} = [-1, -2]\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -d_1 - 2d_2 > 0.$$

从上面各式看出,第一个不等式可化为 $d_1 + d_2 > 0$,而第二个不等式可化为 $d_1 + d_2 < 0$. 所以它们是互相排斥的,亦即不可能存在向量 $\boldsymbol{d} = [d_1, d_2]^T$,使上面三个不等式同时得到满足. 这说明 $\boldsymbol{x}^*$ 满足几何最优性条件. 从图 12－4 也容易看出,在点 $\boldsymbol{x}^*$ 处,下降方向锥 F_0 与可行方向锥 G 的交集是空集.

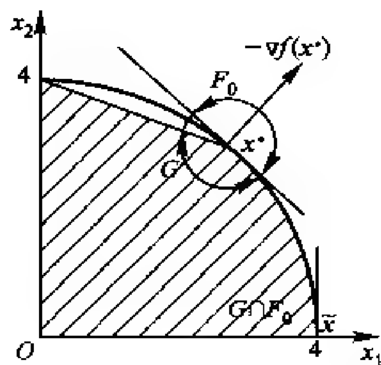


图 12－4

再考查点 $\tilde{x}=[4,0]^T$,由下降可行方向的定义,容易验证,$I(\tilde{x})=\{1,4\}$,且能满足下面三个不等式的可行方向 $d=[d_1,d_2]^T$ 是下降可行方向,

$$\nabla f(\tilde{x})^T d = [0,\ -10]\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -10d_2 < 0,$$

$$\nabla g_1(\tilde{x})^T d = [-2,\ -1]\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -2d_1 - d_2 > 0,$$

$$\nabla g_4(\tilde{x})^T d = [0,1]\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = d_2 > 0,$$

此时,若取 $d_2>0, d_1<-\frac{1}{2}d_2$,那么 d 的方向就是下降可行方向. $\tilde{x}$ 不满足几何最优性条件,自然它肯定不是极小问题的最优点.

12.2.2 充分条件

定理 12.5 若 f 和 $g_i(i\in I(x^*))$ 在点 x^* 处可微,且 f 为凸函数,$g_i(i\in I(x^*))$ 为凹函数,设 x^* 是问题(12-3)的可行点,且有 z 使得

$$\nabla g_i(x^*)^T z > 0, \forall i \in I(x^*).$$

如果在 x^* 点有

$$F_0 \cap G_0 = \varnothing, \tag{12-9}$$

其中 $$G_0 = \{d \mid \nabla g_i(x^*)^T d > 0, i\in I(x^*)\}$$

成立,则 x^* 是该问题的局部极小点.

证 (反证)假定 x^* 不是该问题的局部极小点,同时设 x^0 为该问题的局部极小点,则有

$$f(x^0) < f(x^*), x^0 \neq x^*,$$
$$g_i(x^0) \geqslant 0, i = 1,2,\cdots,m.$$

因为 f 为凸函数,$g_i(i\in I(x^*))$ 为凹函数,所以有

$$0 > f(x^0) - f(x^*) \geqslant \nabla f(x^*)^T(x^0 - x^*),$$
$$0 \leqslant g_i(x^0) - g_i(x^*) \leqslant \nabla g_i(x^*)^T(x^0 - x^*), \forall i \in I(x^*).$$

若记 $y=\lambda z+(1-\lambda)(x^0-x^*)$, $0<\lambda<1$,且 z 满足

$$\nabla g_i(x^*)^T z > 0, \forall i\in I(x^*),$$

则只要 λ 足够小,就有

$$\nabla f(x^*)^T y = \lambda\ \nabla f(x^*)^T z + (1-\lambda)\ \nabla f(x^*)^T(x^0 - x^*) < 0,$$
$$\nabla g_i(x^*)^T y = \lambda\ \nabla g_i(x^*)^T z + (1-\lambda)\ \nabla g_i(x^*)^T(x^0 - x^*) \geqslant$$
$$\lambda\ \nabla g_i(x^*)^T z > 0, i\in I(x^*),$$

这与条件 $F_0 \cap G_0 = \varnothing$ 矛盾，故 x^* 是该问题的局部极小点.

几何最优性条件直观易懂，但是在实际计算中使用起来并不简便. 为此，我们在下节给出 Kuhn - Tucker 条件.

12.3　引用 Lagrange 函数的最优性条件

如前所述，若我们令

$$g(x) = (g_1(x), g_2(x), \cdots, g_m(x))^T,$$
$$h(x) = (h_1(x), h_2(x), \cdots, h_l(x))^T,$$
$$u = (u_1, u_2, \cdots, u_m)^T,$$
$$v = (v_1, v_2, \cdots, v_l)^T.$$

又若令广义 Lagrange 函数为

$$\tilde{L}(x, u, v) = u_0 f(x) - u^T g(x) - v^T h(x),$$

狭义 Lagrange 函数为

$$L(x, u, v) = f(x) - u^T g(x) - v^T h(x),$$

则可定义 F - J 条件，K - T 条件如下：

定义 12.10（F - J 条件）　称如下的条件

$$\begin{cases} L(x^*, u^*, v^*) \leqslant L(x, u^*, v^*), \\ L(x^*, u^*, v^*) = u_0^* f(x^*), \\ (u_0^*, v^*) \geqslant 0, u^*, v^* \geqslant 0, \\ g_i(x^*) \geqslant 0, i = 1, 2, \cdots, m \end{cases} \tag{12-10}$$

为非线性问题(12 - 1)的 F - J 条件，并称满足此条件的点为 F - J 点.

注　以上的 F - J 条件称为非线性化 F - J 条件，特别地，其对应的线性化 F - J 条件为：

$$\begin{aligned} u_0 \nabla f(x^*) - \sum_{i=1}^{m} u_i \nabla g_i(x^*) - \sum_{i=1}^{l} v_i \nabla h_i(x^*) &= 0, \\ u_i g_i(x^*) = 0, i = 1, 2, \cdots, m&, \\ u_i \geqslant 0, i = 0, 1, 2, \cdots, m&. \end{aligned} \tag{12-11}$$

定义 12.11（K - T 条件）　称如下的条件

$$\begin{cases} L(x^*, u^*, v^*) \leqslant L(x, u^*, v^*), \\ L(x^*, u^*, v^*) = f(x^*), \\ u^* \geqslant 0, \\ g_i(x^*) \geqslant 0, i = 1, 2, \cdots, m \end{cases} \tag{12-12}$$

为非线性问题(12－1)的 K－T 条件，并称满足此条件的点为 K－T 点.

注 以上的 K－T 条件称为非线性化 K－T 条件，特别地，其对应的线性化 K－T 条件为：

$$\nabla f(x^*) - \sum_{i=1}^{m} u_i \nabla g_i(x^*) - \sum_{i=1}^{l} v_i \nabla h_i(x^*) = 0,$$
$$u_i g_i(x^*) = 0, i = 1,2,\cdots,m, \tag{12-13}$$
$$u_i \geqslant 0, i = 1,\cdots,m,$$

向量形式为：
$$\begin{cases} \nabla_x L(x^*,u^*,v^*)^{\mathrm{T}}(x - x^*) \geqslant 0 (\forall x \in D), \\ L(x^*,u^*,v^*) = f(x^*), \\ u^* \geqslant 0, \\ g_i(x^*) \geqslant 0, i = 1,2,\cdots,m. \end{cases}$$

定理 12.6 若 $f(x)$，$g_i(x)$ $(i \in I(x^*))$ 是凸函数且于 x^* 处可微，$h_i(x)$ $(i = 1,2,\cdots,l)$，是线性函数，则线性化 K－T 条件与非线性化 K－T 条件等价.

证 "$\Leftarrow$"(反证法)由已知，非线性化 K－T 条件于 (x^*,u^*,v^*) 点满足，即

$$\begin{cases} L(x^*,u^*,v^*) \leqslant L(x,u^*,v^*), \\ L(x^*,u^*,v^*) = f(x^*), \\ u^* \geqslant 0, \\ g_i(x^*) \geqslant 0, i = 1,2,\cdots,m \end{cases}$$

成立，若假定线性化 K－T 条件中，存在 x^0，使得

$$\nabla_x L(x^*,u^*,v^*)^{\mathrm{T}}(x^0 - x^*) < 0 \quad (x^0 \in D),$$

令 $x^0 - x^* = t$，对充分小的 $0 < \lambda < 1$，又令 $x = x^* + \lambda t$，

所以有

$$L(x,u^*,v^*) = L(x^*,u^*,v^*) + \lambda \nabla_x L(x^*,u^*,v^*)^{\mathrm{T}} t + 0(\lambda^2).$$

因为 $\nabla_x L(x^*,u^*,v^*)^{\mathrm{T}} t < 0$，所以对充分小的 $0 < \lambda < 1$ 有

$$L(x,u^*,v^*) < L(x^*,u^*,v^*),$$

且 $x = x^* + \lambda(x^0 - x^*) = \lambda x^0 + (1 - \lambda)x^* \in D$ 与 K－T 条件第一条矛盾，所以一定有

$$\nabla_x L(x^*,u^*,v^*)^{\mathrm{T}}(x - x^*) \geqslant 0 \quad (\forall x \in D).$$

"$\Rightarrow$"若线性化 K－T 条件于 (x^*,u^*,v^*) 点满足，易得 $L(x,u^*,v^*)$ 是凸函数，所以有

$$L(x,u^*,v^*) \geqslant L(x^*,u^*,v^*) + \nabla_x L(x^*,u^*,v^*)^{\mathrm{T}}(x - x^*) \geqslant L(x^*,u^*,v^*) (\forall x \in D),$$

便得到非线性化 K－T 条件.

12.3.1　必要条件

1. Fritz John 条件

不失一般性，下面我们只对线性化情形进行详细推导.

首先，对第一章的定理 1.4(Farkas 引理)我们给出其几何意义. 如图 12－5 是其在二维情形下的几何解释. $\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\alpha}_2$ 张成了凸多面锥 A. 分别与 $\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\alpha}_2$ 成锐角的向量集合是两个半空间(在二维情形下是两个半平面)，这两个半空间的交集是图 12－5 中由向量 $\boldsymbol{d}_1$ 和 $\boldsymbol{d}_2$ 所张成的凸多面锥 B. 显然此锥中的任意向量 $\boldsymbol{d}$ 同时与 $\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\alpha}_2$ 交成锐角.

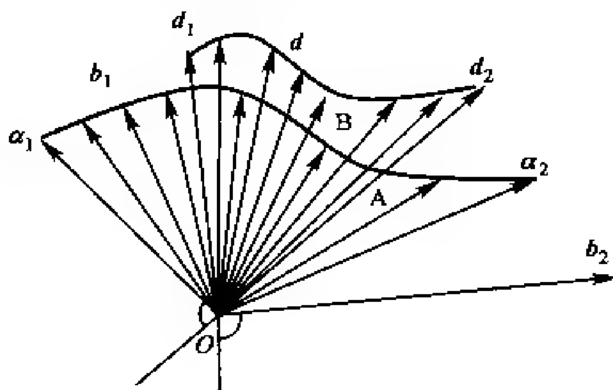


图 12－5

Farkas 引理指出，某一向量与凸多面锥 B 中任意向量都交成锐角的充要条件是，该向量处在凸多面锥 A 之中. 图 12－5 中的 $\boldsymbol{b}_1$ 处在 A 中，它就与 B 中的任意向量交成锐角；反之，$\boldsymbol{b}_2$ 不在 A 中，它就不与 B 中的所有向量交成锐角，例如，$\boldsymbol{d}_1$ 与 $\boldsymbol{b}_2$ 交成钝角.

下面介绍与 Farkas 引理相匹配的另一个引理.

引理 12.1(Gordan)　设 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$ 是 n 维向量，不存在向量 $\boldsymbol{d}$ 使得

$$\boldsymbol{\alpha}_i^T\boldsymbol{d}<0, i=1,2,\cdots,m$$

成立的充要条件是，存在不全为零的非负数 $u_1,u_2,\cdots,u_m$ 使得

$$\sum_{i=1}^{m}u_i\boldsymbol{\alpha}_i=0.$$

该引理也有明显的几何意义. “不存在向量 $\boldsymbol{d}$ 使得 $\boldsymbol{\alpha}_i^T\boldsymbol{d}<0, i=1,2,\cdots,m$”在几何上表示，向量 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$ 不会同时处在过原点 O 的任何超平面 H 的一侧(图 12－6 给出二维的情形). 这时我们总可以适当放大或缩小各向量的长度(相当于选取不全为零的非负数 $u_1,u_2,\cdots,u_m$)对

应去乘 $\alpha_1,\alpha_2,\cdots,\alpha_n$，使得变化后的向量合成为零向量. 这是容易办到的. 例如，在图 12-6 中，若取 $u_1=1,u_2=2,u_3=\frac{1}{2}$，则

$$u_1\alpha_1+u_2\alpha_2+u_3\alpha_3=0.$$

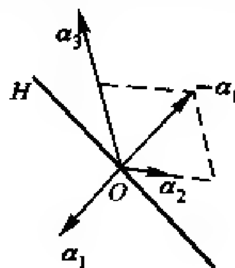


图 12-6

定理 12.7（Fritz John 条件）　设 x^* 是问题（12-1）的局部最优解，又设 $f(x),g_1,g_2,\cdots,g_m,h_1,h_2,\cdots,h_l$ 在点 x^* 处可微，那么必存在不全为零的数

$$u_0,u_1,u_2,\cdots,u_m,v_1,v_2,\cdots,v_l,$$

使得

$$u_0\nabla f(x^*)-\sum_{i=1}^{m}u_i\nabla g_i(x^*)-\sum_{i=1}^{l}v_i\nabla h_i(x^*)=0,$$

$$u_ig_i(x^*)=0,i=1,2,\cdots,m,$$

$$u_i\geqslant 0,i=0,1,2,\cdots,m.$$

证　分两种情形：

（1）若 $\nabla h_1(x^*),\nabla h_2(x^*),\cdots,\nabla h_l(x^*)$ 线性相关，此时存在那么必存在不全为零的数 $v_1,v_2,\cdots,v_l$，使得

$$\sum_{i=1}^{l}v_i\nabla h_i(x^*)=0,$$

所以，只要取 $u_0=0,u_i=0,i=1,2,\cdots,m$，则结论成立.

（2）若 $\nabla h_1(x^*),\nabla h_2(x^*),\cdots,\nabla h_l(x^*)$ 线性无关，设

$$I(x^*)=\{i\mid g_i(x^*)=0,i=1,2,\cdots,m\},$$

因为 x^* 是（12-1）的局部极小点，根据定理 12.4，不存在向量 d 使得

$$\nabla f(x^*)^{\mathrm T}d<0,$$

$$-\nabla g_i(x^*)^{\mathrm T}d<0,i\in I(x^*),$$

$$\nabla h_i(x^*)^{\mathrm T}d=0,i=1,2,\cdots,l$$

同时成立，把 $\nabla f(x^*)$ 和所有的 $-\nabla g_i(x^*)$（其中 $i=1,2,\cdots,m$）依次看作是引理 12.1 中的 $\alpha_1,\alpha_2,\cdots$，那么必存在不全为零的数

$$u_0\geqslant 0,u_i\geqslant 0,i\in I(x^*),v_i(i=1,2,\cdots,l),\qquad(12-14)$$

使得 $$u_0 \nabla f(x^*) - \sum_{i \in I} u_i \nabla g_i(x^*) - \sum_{i=1}^{l} v_i \nabla h_i(x^*) = 0, \tag{12-15}$$

如果对于所有的 $i = 1, 2, \cdots, m$ 但 $i \notin I(x^*)$，令 $u_i = 0$，那么(12-14)和(12-15)可分别写成

$$u_0 \nabla f(x^*) - \sum_{i=1}^{m} u_i \nabla g_i(x^*) - \sum_{i=1}^{l} v_i \nabla h_i(x^*) = 0, \tag{12-16}$$

$$u_i \geqslant 0, i = 0, 1, \cdots, m,$$

且由"对于所有的 $i = 1, 2, \cdots, m$，但 $i \notin I(x^*)$，令 $u_i = 0$"，还可推出

$$u_i g_i(x^*) = 0, i = 1, 2, \cdots, m \tag{12-17}$$

(这是因为，对于 $i \notin I(x^*)$，$g_i(x^*) > 0$，(12-17)就暗示 $u_i = 0$.)把(12-16)和(12-17)，合并起来正是我们要证明的结论.

2. Kuhn-Tucker 条件

我们可看出，当 Fritz John 条件中 $u_0 = 0$ 时，也即 $\nabla f(x^*)$ 的系数为 0，则 $\nabla f(x^*)$ 将从最优性条件中消失，因此最优性条件便仅能说明起作用约束函数的梯度是线性相关的. 而对判别一个可行点是否为极小点就没有了意义，即当 $u_0 = 0$ 时，Fritz-John 条件失效.

分析其失效的原因，我们发现是因为，在所考虑的可行点处，起作用的约束函数的梯度线性相关了. 如果，对起作用的约束函数的梯度附加上线性无关的条件，则可保证 $u_0 > 0$，这便是 Kuhn-Tucker 条件的内容.

定理 12.8(Kuhn-Tucker 条件) 在问题(12-1)中，假设

(1) x^* 是局部最优解；

(2) $f(x), g_1, g_2, \cdots, g_m, h_1, h_2, \cdots, h_l$ 在点 x^* 处可微；

(3) $I(x^*) = \{i \mid g_i(x^*) = 0, i = 1, 2, \cdots, m\}$；

(4) 对于所有 $i \in I$ 的 $\nabla g_i(x^*)$ 与 $\nabla g_1(x^*), \nabla g_2(x^*), \cdots, \nabla g_l(x^*)$ 线性无关.

那么，必存在实数 $u_1, u_2, \cdots, u_m, v_1, v_2, \cdots, v_l$ 使得

$$\nabla f(x^*) - \sum_{i=1}^{m} u_i \nabla g_i(x^*) - \sum_{i=1}^{l} v_i \nabla h_i(x^*) = 0,$$

$$u_i g_i(x^*) = 0, i = 1, 2, \cdots, m,$$

$$u_i \geqslant 0, i = 1, \cdots, m.$$

证 根据 Fritz John 定理，存在不全为零的数 $\tilde{u}_0, \tilde{u}_1, \tilde{u}_2, \cdots, \tilde{u}_m, \tilde{v}_1, \tilde{v}_2, \cdots, \tilde{v}_l$ 使得

$$\tilde{u}_0 \nabla f(x^*) - \sum_{i=1}^{m} \tilde{u}_i \nabla g_i(x^*) - \sum_{i=1}^{l} \tilde{v}_i \nabla h_i(x^*) = 0,$$
$$\tilde{u}_i g_i(x^*) = 0, i = 1,2,\cdots,m, \qquad (12-18)$$
$$\tilde{u}_i \geqslant 0, i = 0,1,2,\cdots,m,$$

由条件(3)知，$\tilde{u}_0 \neq 0$，即 $\tilde{u}_0 > 0$. 若令 $u_i = \dfrac{\tilde{u}_i}{\tilde{u}_0}$，并且用 $\tilde{u}_0$ 除(12－18)的各式两边，则得证.

另外，还有如下定理：

定理 12.9 若 x^* 是(PP)问题的最优解，且 $f(x)$，$g_i(x)(i=1,2,\cdots,m)$是凸函数，且存在 x^0 使得 $g_i(x)>0(i \in I(x^*))$(此条件称为 Slater 条件)成立，则存在 u^*，v^* 使得(x^*, u^*, v^*)为所给问题的 K－T 点(证明略).

例 12.4 求问题

$$\min f(x) = x_1,$$
$$g_1(x) = 16 - (x_1 - 4)^2 - x_2^2 \geqslant 0,$$
$$h_1(x) = (x_1 - 3)^2 + (x_2 - 2)^2 - 13 = 0$$

的 K－T 点.

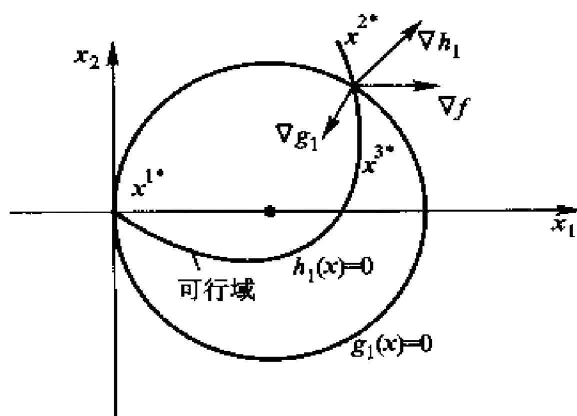


图 12－7

解 如图 12－7 所示，问题的可行域 D 为圆周 $h_1(x)=0$ 在圆 $g_1(x)=0$内的那部分. 因为对任意的 $x \in D$，都有 $\nabla f(x) = (1,0)^T$.

又知，

$$\nabla g_1(x) = [-2(x_1-4), -2x_2]^T, \nabla h_1(x)$$
$$= [2(x_1-3), 2(x_2-2)]^T,$$

而对于以下三组解

$$\begin{cases} x^{1*} = (0,0)^{\mathrm{T}}, \\ u^{1*} = \dfrac{1}{8}, \\ v^{1*} = 0. \end{cases} \quad \begin{cases} x^{2*} = (6.4,3.2)^{\mathrm{T}}, \\ u^{2*} = \dfrac{3}{40}, \\ v^{2*} = \dfrac{1}{5}. \end{cases} \quad \begin{cases} x^{3*} = (3+\sqrt{13},2)^{\mathrm{T}}, \\ u^{3*} = 0, \\ v^{3*} = \dfrac{\sqrt{13}}{26}. \end{cases}$$

易验证都满足 K－T 条件，即 x^{1*}, x^{2*}, x^{3*} 均为 K－T 点，但由图上不难看出，只有 x^{2*}, x^{3*} 才是问题的局部极小解. 所以 K－T 条件是局部极小的必要非充分条件.

12.3.2　充分条件

定理 12.10　设 x^* 是如下凸规划问题：

$$\begin{cases} \min f(x), \\ \text{s.t.} \quad g_i(x) \geqslant 0, i = 1,2,\cdots,m, \end{cases} \tag{12-19}$$

其中，f 为凸函数，$g_i(x) \geqslant 0, i = 1,2,\cdots,m$ 为凹函数的一个可行解，且在 x^* 处 K－T 条件成立，则 x^* 是此凸规划问题的最优解.

证　因为 $g_i(x) \geqslant 0, i = 1,2,\cdots,m$ 为凹函数，所以有

$$0 \leqslant g_i(x) \leqslant g_i(x^*) + \nabla g_i(x^*)^{\mathrm{T}}(x - x^*)$$
$$= \nabla g_i(x^*)^{\mathrm{T}}(x - x^*), i \in I(x^*).$$

又因为，在 x^* 处 K－T 条件成立，即

$$\nabla f(x^*) - \sum_{i=1}^{m} u_i \nabla g_i(x^*) = 0,$$

所以

$$\nabla f(x^*) = \sum_{i=1}^{m} u_i \nabla g_i(x^*),$$

因此得

$$\nabla f(x^*)^{\mathrm{T}}(x - x^*) = \sum_{i \in I(x^*)}^{m} u_i \nabla g_i(x^*)^{\mathrm{T}}(x - x^*) \geqslant 0,$$

而 $f(x)$ 为凸函数，故对任意 $x \in D$ 有

$$f(x) \geqslant f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*) \geqslant f(x^*),$$

所以，x^* 是此凸规划问题的最优解.

而且在凸规划部分我们已经证明了，凸规划问题的任何局部极小点都是全局极小点，所以此处的最优解亦是全局最优解.

定理 12.11　若 (x^*, u^*, v^*) 是(PP)问题满足非线性化 K－T 条件的点，则 x^* 为问题的最优点.

证 因为 $u^* \geqslant 0$,所以对 $x \in D$ 都有

$$f(x^*) = L(x^*, u^*, v^*) \leqslant L(x, u^*, v^*)$$
$$= f(x) - \sum_{i=1}^{m} u_i^* \nabla g_i(x) - \sum_{i=1}^{l} v_i^* \nabla h_i(x)$$
$$= f(x) - \sum_{i=1}^{m} u_i^* \nabla g_i(x) \leqslant f(x),$$

所以,x^* 是(PP)问题的最优解.

定理 12.12 设 (1) x^* 为非线性问题(12-1)的一个可行解;

(2) 若存在 $u_1, u_2, \cdots, u_m, v_1, v_2, \cdots, v_l (u_i \geqslant 0, i = 1, \cdots, m)$ 使 K-T 条件成立;

(3) 对任何满足

$$p^{\mathrm{T}} \nabla h_i(x^*) = 0, i = 1, 2, \cdots, l,$$

$$p^{\mathrm{T}} \nabla g_i(x^*) \geqslant 0, i \notin I(x^*) \text{ 且 } u_i^* = 0, i = 1, 2, \cdots, m,$$

$$p^{\mathrm{T}} \nabla g_i(x^*) = 0, i \in I(x^*) \text{ 且 } u_i^* > 0, i = 1, 2, \cdots, m$$

的向量 $p \neq 0$,都有

$$p^{\mathrm{T}} \nabla_{xx}^2 L(x^*, u^*, v^*) p > 0,$$

则 x^* 为非线性问题(12-1)的严格局部极小点.(证明略)

综合前面的讨论,我们可得出如下的最优性条件:

1. 若(x^*, u^*, v^*)是(PP)问题满足非线性化 K-T 条件的点,则 x^* 为(PP)问题的最优点. 反之,若 x^* 为(PP)问题的最优点,且 $f(x), g_i(x)$ $(i = 1, 2, \cdots, m)$ 是凸函数,且存在 x^0 使得 $g_i(x) > 0 (i \in I(x^*))$(此条件称为 Slater 条件)成立,则存在 u^*, v^* 使得(x^*, u^*, v^*)为所给问题的 K-T 点.

2. 若 x^* 为(PP)问题的最优点,则存在 u^*, v^* 使得(x^*, u^*, v^*)为(PP)问题的线性化 K-T 点. 反之,若(x^*, u^*, v^*)为(PP)问题的线性化 K-T 点,且 $f(x), g_i(x) (i \in I(x^*))$ 是凸函数且于 x^* 处可微,则 x^* 为(PP)问题的最优点.

例 12.5 试分析求解下列问题

$$\begin{cases} \min & f(x) = x_1^2 + x_2, \\ \text{s.t.} & g_1(x) = -(x_1 + x_2^2) + 1 \geqslant 0, \\ & g_2(x) = -(x_1 + x_2) + 1 \geqslant 0, \\ & h(x) = x_1^2 + x_2^2 - 9 = 0. \end{cases}$$

解 求解 $g_1(x)=0, h(x)=0$ 的交点，得其坐标为 $x^*=(-2.37, -1.84)^T$.

将其代入 $g_2(x)$ 得：$g_2(x^*)=-(-2.37-1.84)+1=5.21>0$，所以满足约束条件，且 $g_1(x)\geqslant 0, h(x)=0$ 为 x^* 处起作用的约束.

1. 检查在 x^* 处起作用的约束的梯度是否线性相关.

令
$$c_1 \nabla g_1(x^*) + c_2 \nabla h(x^*) = 0,$$

即
$$c_1\begin{bmatrix}-1\\-2x_2^*\end{bmatrix}+c_2\begin{bmatrix}2x_1^*\\2x_2^*\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix},$$

可知，只有当 $c_1=c_2=0$ 时上式才成立，故 $\nabla g_1(x^*), \nabla h(x^*)$ 线性无关.

2. 检查 K－T 条件是否满足

$$\nabla f(x^*) - u_1^* \nabla g_1(x^*) - u_2^* \nabla g_2(x^*) - v^* \nabla h(x^*) = 0,$$

即
$$\begin{cases}\begin{bmatrix}2x_1^*\\1\end{bmatrix}-u_1^*\begin{bmatrix}-1\\-2x_2^*\end{bmatrix}-u_2^*\begin{bmatrix}-1\\-1\end{bmatrix}-v^*\begin{bmatrix}2x_1^*\\2x_2^*\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix},\\ u_1^*(1-x_1^*-x_2^{*2})=0,\\ u_2^*(1-x_1^*-x_2^*)=0,\\ u_1^*\geqslant 0, u_2^*\geqslant 0,\end{cases}$$

解得 $u_1^*=0, u_2^*=1.05, v^*=0.778$，满足 K－T 条件.

3. 检查充分条件：

$$\nabla^2 f(x^*)=\begin{bmatrix}2&0\\0&0\end{bmatrix}, \nabla^2 h(x^*)=\begin{bmatrix}2&0\\0&2\end{bmatrix},$$

$$\nabla^2 g_1(x^*)=\begin{bmatrix}0&0\\0&-2\end{bmatrix}, \nabla^2 g_2(x^*)=\begin{bmatrix}0&0\\0&0\end{bmatrix},$$

$$[P_1\quad P_2]\left\{\begin{bmatrix}2&0\\0&0\end{bmatrix}-v^*\begin{bmatrix}2&0\\0&2\end{bmatrix}-u_2^*\begin{bmatrix}0&0\\0&-2\end{bmatrix}\right\}\begin{bmatrix}P_1\\P_2\end{bmatrix}$$

$$=[P_1\quad P_2]\left\{\begin{bmatrix}2&0\\0&0\end{bmatrix}-0.778\begin{bmatrix}2&0\\0&2\end{bmatrix}-1.05\begin{bmatrix}0&0\\0&-2\end{bmatrix}\right\}\begin{bmatrix}P_1\\P_2\end{bmatrix}$$

$$=[P_1\quad P_2]\begin{bmatrix}0.444&0\\0&0.544\end{bmatrix}\begin{bmatrix}P_1\\P_2\end{bmatrix}$$

$$=0.444P_1^2+0.544P_2^2>0,$$

即对 $P\neq 0$ 均成立，所以 $x^*=(-2.37, -1.84)^T$ 为严格极小点.

第 13 章 非线性规划的对偶理论

本书第 11 章中介绍了线性规划的对偶理论，本章所介绍的非线性规划的对偶理论是对前述内容的进一步扩充，它对于约束最优化问题的最优解的性质及最优性条件的揭示等都起着重要的作用.

给出一个非线性规划问题(我们称之为原问题)，与它有密切联系的是另一个非线性规划问题(后者称为对偶问题)，在某种凸性的假设下，原问题与对偶问题的最优解之间以及最优目标值之间有密切关系，因而原问题的求解可以间接地归结为对偶问题的求解.

13.1 Lagrange 对偶问题与弱对偶性定理

设原问题为：

$$
\begin{aligned}
(\mathrm{PP}) \quad & \min \quad f(\boldsymbol{x}), \boldsymbol{x} \in \mathbf{R}^n, \\
& \text{s.t.} \quad \boldsymbol{g}(\boldsymbol{x}) \geqslant 0, \\
& \qquad\ \ \boldsymbol{h}(\boldsymbol{x}) = 0, \boldsymbol{x} \in X,
\end{aligned}
$$

其中 $\boldsymbol{g}(\boldsymbol{x})=(g_1(\boldsymbol{x}),g_2(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x}))^{\mathrm{T}}$，$\boldsymbol{h}(\boldsymbol{x})=(h_1(\boldsymbol{x}),h_2(\boldsymbol{x}),\cdots,h_l(\boldsymbol{x}))^{\mathrm{T}}$，$f(\boldsymbol{x})$，$g_i(\boldsymbol{x})(i=1,2,\cdots,m)$，$h_i(\boldsymbol{x})(i=1,2,\cdots,l)$ 均为定义在 $X \subset \mathbf{R}^n$ 上的实值函数. 下面我们考虑原问题(PP)的对偶规划问题：

$$
\begin{aligned}
(\mathrm{DP}) \quad & \max \quad \varphi(\boldsymbol{u},\boldsymbol{v}), \\
& \text{s.t.} \quad \boldsymbol{u} \geqslant 0,
\end{aligned}
$$

其中 $\varphi(\boldsymbol{u},\boldsymbol{v})=\min\limits_{\boldsymbol{x}\in X} L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})$，$L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})$ 是原问题的 Lagrange 乘子函数，即

$$
\begin{aligned}
L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}) &= f(\boldsymbol{x}) - \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}) \\
&= f(\boldsymbol{x}) - \sum_{i=1}^{m} u_i g_i(\boldsymbol{x}) - \sum_{i=1}^{l} v_i h_i(\boldsymbol{x}).
\end{aligned}
$$

在对偶问题(DP)中的目标函数 $\varphi(\boldsymbol{u},\boldsymbol{v})$，我们称之为 Lagrange 对偶函数，其中与不等式约束 $g_i(\boldsymbol{x})\geqslant 0$ 相匹配的乘子 u_i 是非负的，而与等式约束 $h_i(\boldsymbol{x})=0$ 匹配的乘子 v_i 不受符号的限制.

由于对偶问题是极大化函数 $L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})$ 的极小值组成，因此又称(DP)为极大 - 极小对偶问题.

对于原问题(PP)，可以作出不同的对偶问题(DP)，比如当 $f(\boldsymbol{x})$，

$g_i(\boldsymbol{x}),h_i(\boldsymbol{x})$是在 $X\subset\mathbf{R}^n$ 上具有一阶连续偏导数且为凸函数时，可以考虑如下面的对偶问题：

$$\begin{aligned}(\text{DP1})\quad \max\quad & L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}),\\ \text{s.t.}\quad & L_x(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})=0,\\ & \boldsymbol{u}\geqslant 0,\end{aligned}$$

亦即

$$\begin{aligned}\max\quad & L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}),\\ \text{s.t.}\quad & f_x(\boldsymbol{x})-\boldsymbol{u}^{\mathrm{T}}g_x(\boldsymbol{x})-\boldsymbol{v}^{\mathrm{T}}h_x(\boldsymbol{x})=0,\\ & \boldsymbol{u}\geqslant 0.\end{aligned}$$

所谓非线性规划的对偶问题理论，就是研究原问题(PP)与对偶问题(DP)或(DP1)之间的关系. 它包括(PP)与(DP)的最优目标函数值之间的关系以及(PP)与(DP)的最优解之间的关系.

例 13.1 设原问题

$$\begin{aligned}\min\quad & x_1^2+x_2^2,\\ \text{s.t.}\quad & -x_1-x_2+4\leqslant 0,\\ & x_1,x_2\geqslant 0,\end{aligned}$$

原问题的最优解 $\boldsymbol{x}^*=(2,2)^{\mathrm{T}}$，对应的最优目标值等于 8.

令 $g(\boldsymbol{x})=x_1+x_2-4,\boldsymbol{X}=\{(x_1,x_2)\mid x_1,x_2\geqslant 0\}$，

对偶函数为

$$\begin{aligned}\varphi(u)&=\min\{x_1^2+x_2^2-u(x_1+x_2-4)\mid x_1,x_2\geqslant 0\}\\ &=\min\{x_1^2-ux_1\mid x_1\geqslant 0\}+\min\{x_2^2-ux_2\mid x_2\geqslant 0\}+4u,\end{aligned}\tag{13-1}$$

显然，如果 $u\geqslant 0$，式(13－1)中的极小值在 $x_1=x_2=\dfrac{u}{2}$处达到；如果 $u<0$，式(13－1)中的极小值在 $x_1=x_2=0$ 处达到，因此

$$\varphi(u)=\begin{cases}-\dfrac{1}{2}u^2+4u, & u\geqslant 0,\\ 4u, & u<0,\end{cases}$$

对偶函数 $\varphi(u)$为凸函数，并且当 $u\geqslant 0$ 时它在 $u^*=4$ 取得极大值 8. 可见原问题与对偶问题二者的最优目标值相等.

下面的对偶定理及其推论指出，对于可行点而言，原问题与对偶问题的目标值不必相等，但若二者相等时，则原问题的可行解与对偶问题的可行解将分别成为最优解.

定理 13.1(弱对偶定理) 设 $f(\boldsymbol{x}),g_i(\boldsymbol{x})(i=1,2,\cdots,m),h_i(\boldsymbol{x})(i=1,2,\cdots,l)$均为定义在 $X\subset\mathbf{R}^n$ 上的实值函数，则对于(PP)问题的可行解 $\bar{\boldsymbol{x}}$ 及(DP)问题的可行解$(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$均有

$$f(\bar{x}) \geqslant \varphi(\bar{u},\bar{v}). \qquad (13-2)$$

证　由定义,对任意的可行解 $\bar{x}$,$(\bar{u},\bar{v})$,$\bar{u}\geqslant 0$,均有

$$\begin{aligned}\varphi(\bar{u},\bar{v}) &= \min\{f(x) - \bar{u}^{T}g(x) - \bar{v}^{T}h(x) \mid x \in X\}\\ &\leqslant f(\bar{x}) - \bar{u}^{T}g(\bar{x}) - \bar{v}^{T}h(\bar{x})\\ &\leqslant f(\bar{x}),\end{aligned}$$

即式(13-2)成立.

推论 13.1　若 $f(x)$,$g_i(x)(i=1,2,\cdots,m)$,$h_i(x)(i=1,2,\cdots,l)$为定义在 $X\subset\mathbf{R}^n$ 上的实值函数,$\bar{x}$、$(\bar{u},\bar{v})$分别为(PP)及(DP)的可行解,且满足

$$f(\bar{x}) = \varphi(\bar{u},\bar{v}),$$

则 $\bar{x}$,$(\bar{u},\bar{v})$分别为(PP)及(DP)的最优解.

证　据假设 $\bar{x}$、$(\bar{u},\bar{v})$分别为(PP)及(DP)的一个可行解,又设 x 为(PP)的任意可行解,由定理 13.1 及本推论的假设,总有

$$f(x) \geqslant \varphi(\bar{u},\bar{v}) = f(\bar{x}),$$

上式表明 $\bar{x}$ 为(PP)的最优解.

若设(u,v)分别为(DP)的任一个可行解,同理得

$$\varphi(u,v) \leqslant f(\bar{x}) = \varphi(\bar{u},\bar{v}),$$

可见,$(\bar{u},\bar{v})$亦为(DP)问题的最优解.

例 13.2　设原问题(PP)为:

$$\begin{aligned}\min\quad & f(x) = x_1^2 - x_2,\\ \text{s.t.}\quad & g_1(x) = x_1 - 1 \geqslant 0,\\ & g_2(x) = -x_1^2 - x_2^2 + 26 \geqslant 0,\\ & h(x) = x_1 + x_2 - 6 = 0,\end{aligned}$$

其对偶问题(DP)为:

$$\begin{aligned}\max\quad & \varphi(u,v),\\ \text{s.t.}\quad & u \geqslant 0.\end{aligned}$$

其中 $\varphi(u,v)=\min\{x_1^2 - x_2 + u_1(x_1-1) + u_2(26 - x_1^2 - x_2^2) + v(x_1 + x_2 - 6)\}$.

若给出 $\bar{x}=(1,5)^T$,$\bar{u}=(0,\frac{3}{8})^T$,$\bar{v}=\frac{11}{4}$,容易验证,$\bar{x}$ 为(PP)的可行解,而$(\bar{u},\bar{v})$为(DP)的可行解,并且

$$f(\bar{x}) = \varphi(\bar{u},\bar{v}) = 24,$$

据推论 13.1 知,$\bar{x}$、$(\bar{u},\bar{v})$分别为(PP)及(DP)的最优解. 读者还可根据第 12 章中的定理 12.9 验证 $\bar{x}$ 为原问题(PP)的最优解.

13.2 鞍点判别条件

定义 13.1(鞍点) 对于原问题(PP),设 Lagrange 乘子函数

$$L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}) = f(\boldsymbol{x}) - \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}),$$

其中 $\boldsymbol{x} \in \mathbf{R}^n, \boldsymbol{u} \in \mathbf{R}^m, \boldsymbol{v} \in \mathbf{R}^l$,

称点$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$,$\bar{\boldsymbol{u}}\geqslant 0$ 为 $L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})$的鞍点,若对任意的 $\boldsymbol{x}$,$\boldsymbol{u}(\boldsymbol{u}\geqslant 0)$及 $\boldsymbol{v}$,总有下式成立:

$$L(\bar{\boldsymbol{x}},\boldsymbol{u},\boldsymbol{v}) \leqslant L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}) \leqslant L(\boldsymbol{x},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}), \qquad (13-3)$$

即是说关于 $\boldsymbol{x}$,$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$是 L 的极小点,而关于$(\boldsymbol{u},\boldsymbol{v})$,$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$是 L 的极大点. 给定原问题(PP),求$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$满足式(13-3)的问题又称为鞍点问题.

关于鞍点与 K-T 点和最优解之间的关系有如下结论.

定理 13.2(鞍点判别条件 1) 若$f(\boldsymbol{x})$,$g_i(\boldsymbol{x})(i=1,2,\cdots,m)$,$h_i(\boldsymbol{x})(i=1,2,\cdots,l)$均为定义在 $X\subset\mathbf{R}^n$ 上的实值函数,点$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$为 $L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})$的鞍点的充分必要条件为$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$为原问题(PP)的非线性 K-T 点.

证 不失一般性,在(PP)中我们只考虑不等式约束.

首先证充分性. 设在点$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$,(PP)的非线性 K-T 条件被满足,即

(1) 对任意的可行点 $\boldsymbol{x} \in X$,有 $L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})\leqslant L(\boldsymbol{x},\bar{\boldsymbol{u}})$,

(2) $L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}}) = f(\bar{\boldsymbol{x}})$,

(3) $\bar{\boldsymbol{u}}\geqslant 0$,

(4) $g_i(\bar{\boldsymbol{x}})\geqslant 0, i=1,2,\cdots,m$.

由条件(1),对任何可行点 $\boldsymbol{x}$,总有

$$L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}}) \leqslant L(\boldsymbol{x},\bar{\boldsymbol{u}}), \qquad (13-4)$$

又由条件(2),(3),(4),对任何 $\bar{\boldsymbol{u}}\geqslant 0$,有

$$L(\bar{\boldsymbol{x}},\boldsymbol{u}) = f(\bar{\boldsymbol{x}}) - \sum_{i=1}^{m} u_i g_i(\bar{\boldsymbol{x}}) \leqslant f(\bar{\boldsymbol{x}}) = L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}}), \qquad (13-5)$$

综合(13-4),(13-5)有

$$L(\bar{\boldsymbol{x}},\boldsymbol{u}) \leqslant L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}}) \leqslant L(\boldsymbol{x},\bar{\boldsymbol{u}}),$$

即$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$为 $L(\boldsymbol{x},\boldsymbol{u})$的鞍点.

下证必要性,设$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$为 $L(\boldsymbol{x},\boldsymbol{u})$的鞍点,据鞍点的定义,显然 K-T 条件(1),(3)自然满足. 下面用反证法证明 K-T 条件(4)成立.

设相反,若存在 $i_0(1\leqslant i_0\leqslant m)$使 $g_{i_0}(\bar{\boldsymbol{x}})<0$,我们取 $u_i=\bar{u}_i(i\neq i_0)$,$u_{i_0}<\bar{u}_0$,则有

$$L(\bar{x},u)-L(\bar{x},\bar{u})=\sum_{i=1}^{m}(\bar{u}_i-u_i)g_i(\bar{x})=(\bar{u}_{i_0}-u_{i_0})g_{i_0}(\bar{x})>0,$$

即对任意的 $u\geqslant 0$，$L(\bar{x},u)>L(\bar{x},\bar{u})$，这与 $(\bar{x},\bar{u})$ 为 $L(x,u)$ 的鞍点矛盾，所以 $g_i(\bar{x})\geqslant 0, i=1,2,\cdots,m$，即 K－T 条件(4)成立.

欲证明 K－T 条件(2)成立，即 $f(\bar{x})=L(\bar{x},\bar{u})$，即只须证明 $\sum_{i=1}^{m}\bar{u}_i g_i(\bar{x})=0$，亦即

$$\bar{u}_i g_i(\bar{x})=0(i=1,\cdots,m).$$

设相反，若有 $\bar{u}_{i_0}g_{i_0}(\bar{x})>0$，则必有 $\bar{u}_{i_0}>0$ 及 $g_{i_0}(\bar{x})>0$，我们取 $u_i=\bar{u}_i(i\neq i_0)$，$u_{i_0}=0$，则有

$$L(\bar{x},u)-L(\bar{x},\bar{u})=\sum_{i=1}^{m}(\bar{u}_i-u_i)g_i(\bar{x})=\bar{u}_{i_0}g_{i_0}(\bar{x})>0,$$

即对任意的 $u\geqslant 0$，$L(\bar{x},u)>L(\bar{x},\bar{u})$，这与 $(\bar{x},\bar{u})$ 为 $L(x,u)$ 的鞍点矛盾，所以

$$\bar{u}_i g_i(\bar{x})=0, i=1,\cdots,m,$$

亦即

$$\sum_{i=1}^{m}\bar{u}_i g_i(\bar{x})=0.$$

推论 13.2 若 $(\bar{x},\bar{u},\bar{v})$ 是 $L(x,u,v)$ 的鞍点，则 $\bar{x}$ 是原问题(PP)的最优解.

证 据定理 13.2，$(\bar{x},\bar{u},\bar{v})$ 是(PP)的非线性 K－T 点，另据第 12 章定理 12，$\bar{x}$ 是(PP)的最优解.

定理 13.3（鞍点判别条件 2——对偶定理）

若 $f(x)$，$g_i(x)(i=1,2,\cdots,m)$，$h_i(x)(i=1,2,\cdots,l)$ 均为定义在 $X\subset\mathbf{R}^n$ 上的实值函数，点 $(\bar{x},\bar{u},\bar{v})$ 为 $L(x,u,v)$ 的鞍点的充分必要条件为

(1) $\bar{x}$ 是(PP)的最优解；

(2) $(\bar{u},\bar{v})$ 是(DP)的最优解；

(3) $\varphi(\bar{u},\bar{v})=f(\bar{x})$.

证 必要性. 由推论 13.2 可知，$\bar{x}$ 是(PP)的最优解，从而结论(1)成立；

由 $\varphi(\bar{u},\bar{v})$ 的定义，有

$$\varphi(\bar{u},\bar{v})=\min_{x\in X}\{f(x)-\bar{u}^{\mathrm{T}}g(x)-\bar{v}^{\mathrm{T}}h(x)\}=\min_{x\in X}L(x,\bar{u},\bar{v}),$$

因为 $(\bar{x},\bar{u},\bar{v})$ 是 $L(x,u,v)$ 的鞍点，即对任意 $x\in X\subset\mathbf{R}^n$ 有

$$L(\bar{x},\bar{u},\bar{v})\leqslant L(x,\bar{u},\bar{v}),$$

故

$$\varphi(\bar{u},\bar{v})=L(\bar{x},\bar{u},\bar{v})=f(\bar{x})-\bar{u}^{\mathrm{T}}g(\bar{x})-\bar{v}^{\mathrm{T}}h(\bar{x}),$$

由结论(1)，$h(\bar{x})=0$，
又由定理 13.2 中证明，必有

$$\bar{u}^{\mathrm{T}}g(\bar{x})=0,$$

于是得到

$$\varphi(\bar{u},\bar{v})=f(\bar{x}),$$

即结论(3)成立. 又由推论 13.1 可知，$(\bar{u},\bar{v})$ 是(DP)的最优解，即结论(2)亦成立.

充分性. 据结论(1)，(2)，(3)，读者不难证明，对于 $\bar{x},\bar{u}$ $(\bar{u}\geqslant 0)$，$\bar{v}$，以及任意的 $x\in X\subset\mathbf{R}^n$，任意的 $u\geqslant 0$，v，下列不等式成立.

$$f(\bar{x})-\bar{u}^{\mathrm{T}}g(\bar{x})-\bar{v}^{\mathrm{T}}h(\bar{x})\leqslant f(x)-\bar{u}^{\mathrm{T}}g(x)-\bar{v}^{\mathrm{T}}h(x),$$
$$f(\bar{x})-u^{\mathrm{T}}g(\bar{x})-v^{\mathrm{T}}h(\bar{x})\leqslant f(\bar{x})-\bar{u}^{\mathrm{T}}g(\bar{x})-\bar{v}^{\mathrm{T}}h(\bar{x}),$$

亦即下式成立：

$$L(\bar{x},u,v)\leqslant L(\bar{x},\bar{u},\bar{v})\leqslant L(x,\bar{u},\bar{v}),$$

点 $(\bar{x},\bar{u},\bar{v})$ 是 $L(x,u,v)$ 的鞍点.

例 13.3　设原问题(PP)为

$$\begin{aligned}\min\quad & f(x)=x_1^2-x_2,\\ \text{s.t.}\quad & g_1(x)=x_1-1\geqslant 0,\\ & g_2(x)=-x_1^2-x_2^2+26\geqslant 0,\\ & h(x)=x_1+x_2-6=0,\end{aligned}$$

对应的 Lagrange 函数为

$$\begin{aligned}L(x,u,v)=\;&x_1^2-x_2-u_1(x_1-1)-u_2(26-x_1^2-x_2^2)-\\ &v(x_1+x_2-6),u_1\geqslant 0,u_2\geqslant 0,\end{aligned}$$

在例 13.2 中，我们根据推论 13.1 证明了 $\bar{x}=(1,5)^{\mathrm{T}}$ 为(PP)问题的最优解；$\bar{u}=\left(0,\dfrac{3}{8}\right)^{\mathrm{T}}$，$\bar{v}=\dfrac{11}{4}$ 为对偶问题(DP)的最优解，并且二者的目标函数值相等. 因此，根据定理 13.2 的充分条件可以判定，$(\bar{x},\bar{u},\bar{v})$ 应为 $L(x,u,v)$ 的鞍点.

事实上我们可以根据鞍点的定义直接验证，$(\bar{x},\bar{u},\bar{v})$ 是 $L(x,u,v)$ 的鞍点. 因为

$$L(\bar{x},u,v)=-4=L(\bar{x},\bar{u},\bar{v}),\tag{13-6}$$

而　$L(x,\bar{u},\bar{v})=x_1^2-x_2-\dfrac{3}{8}(26-x_1^2-x_2^2)-\dfrac{11}{4}(x_1+x_2-6)$，

又因为

$$\begin{cases}\dfrac{\partial L}{\partial x_1}=\dfrac{11}{4}x_1-\dfrac{11}{4}=0,\\[2ex] \dfrac{\partial L}{\partial x_2}=\dfrac{3}{4}x_2-\dfrac{15}{4}=0,\end{cases}$$

解得驻点 $x_1=1, x_2=5$，且 $\nabla_x^2 L=\begin{bmatrix}\frac{11}{4} & 0\\ 0 & \frac{3}{4}\end{bmatrix}>0$，

故 $\bar{x}=(1,5)^T$ 为 $L(x,\bar{u},\bar{v})$ 的极小点，即

$$L(\bar{x},\bar{u},\bar{v}) \leqslant L(x,\bar{u},\bar{v}). \tag{13-7}$$

综合(13－6)，(13－7)可得

$$L(\bar{x},u,v) = L(\bar{x},\bar{u},\bar{v}) \leqslant L(x,\bar{u},\bar{v}),$$

所以，$(\bar{x},\bar{u},\bar{v})$ 为 $L(x,u,v)$ 的鞍点.

定理 13.2 及其推论和定理 13.3 从理论上揭示了鞍点，K－T 点和最优点三者之间的关系. 即 Lagrange 函数的鞍点和原问题(PP)的 K－T 点之间具有等价关系；原问题(PP)连同其对偶问题(DP)的最优点与 Lagrange 函数的鞍点具有等价关系. 因此据定理 13.2，可直接判定 K－T 点为鞍点，但若据定理 13.3 不能单独由(PP)的最优点或(DP)的最优点判定 Lagrange 函数的鞍点的存在.

下面的定理表明，若在原问题(PP)中，对目标函数 $f(x)$ 和约束函数 $-g(x)$，$h(x)$ 增加凸性的假设，则可以得到一个由最优点直接判定鞍点存在的充分条件.

定理 13.4　设

(1) $f(x)$，$-g_i(x)\ (i=1,2,\cdots,m)$ 均为定义在 $X\subset \mathbf{R}^n$ 上的实值函数，$h_i(x)\ (i=1,2,\cdots,l)$ 是 $X\subset \mathbf{R}^n$ 上的线性函数；

(2) $g_i(x)\ (i=1,2,\cdots,m)$ 满足 Slater 约束品格，即存在 x^0，使 $g_i(x^0)>0\ (i=1,2,\cdots,m)$；

(3) 原问题(PP)存在最优解 $\bar{x}$，则必存在 $\bar{u}>0$，$\bar{v}$ 使 $(\bar{x},\bar{u},\bar{v})$ 是 $L(x,u,v)$ 的鞍点.

这个定理可以作为一个练习，建议读者给予证明.(提示：利用 K－T 条件和 Farkas 引理)

13.3　扩展的对偶定理

本节将引入强对偶定理进一步揭示原问题(PP)和对偶问题(DP)二者存在最优解之间的关系. 在一些加强条件(如凸性条件，可微条件等)下，还可以扩展弱对偶定理，对偶定理以及强对偶定理.

定理 13.5(强对偶定理)　设(1) $f(x)$，$-g_i(x)\ (i=1,2,\cdots,m)$ 均为定义在 $X\subset \mathbf{R}^n$ 上的实值函数，$h_i(x)\ (i=1,2,\cdots,l)$ 是 $X\subset \mathbf{R}^n$ 上的线性

函数；

(2) $g_i(x)(i=1,2,\cdots,m)$满足 Slater 条件；

(3) 若(PP)存在最优解 $\bar{x}$,

则对偶问题亦存在最优解$(\bar{u},\bar{v})$,并且$f(\bar{x})=\varphi(\bar{u},\bar{v})$.

证　在假设条件下,由定理 13.4 可知,存在 $\bar{u}>0,\bar{v}$ 使$(\bar{x},\bar{u},\bar{v})$是$L(x,u,v)$的鞍点,又由定理 13.3 可知$(\bar{u},\bar{v})$是对偶问题(DP)的最优解,并且$f(\bar{x})=\varphi(\bar{u},\bar{v})$.

下面我们考虑原问题(PP)与另一种对偶问题(DP1).不失一般性,我们假定只考虑不等式约束,原问题(PP)为

$$\begin{aligned} \min\quad & f(x), \\ \text{s.t.}\quad & g_i(x)\geqslant 0(i=1,2,\cdots,m), \\ & x\in X, \end{aligned}$$

对偶问题(DP)为

$$\begin{aligned} \max\quad & \varphi(u), \\ \text{s.t.}\quad & u\geqslant 0, \end{aligned}$$

其中

$$\begin{aligned} \varphi(u) &= \min_{x\in X} L(x,u), \\ &= \min_{x\in X}\left\{f(x)-\sum_{i=1}^{m}u_i g_i(x)\right\}. \end{aligned}$$

如果我们对$f(x),g(x)$添加上凸性和可微条件,即假设$f(x),-g_i(x)$ $(i=1,2,\cdots,m)$为$X\subset\mathbf{R}^n$上具有一阶连续偏导数的凸函数,则对偶问题变为如下的非线性规划：

(DP1)

$$\begin{aligned} \max\quad & L(x,u), \\ \text{s.t.}\quad & L_x(x,u)=0, \\ & u\geqslant 0, \end{aligned}$$

亦即

$$\begin{aligned} \max\quad & \{f(x)-u^{\mathrm{T}}g(x)\}, \\ \text{s.t.}\quad & f_x(x)-u^{\mathrm{T}}g_x(x)=0, \\ & u\geqslant 0. \end{aligned}$$

读者不难验证,对于原问题(PP)和对偶问题(DP1)相应的弱对偶定理及其推论(定理 13.1,推论 13.1),对偶定理(定理 13.3),强对偶定理(定理 13.5)均成立.现分别叙述于下：

定理 13.6（弱对偶定理） 设 $f(\boldsymbol{x})$，$-g_i(\boldsymbol{x})(i=1,2,\cdots,m)$ 均为 $X\subset\mathbf{R}^n$ 上具有一阶连续偏导数的凸函数，则对于原问题（PP）的可行解 $\bar{\boldsymbol{x}}$，及对偶问题（DP1）的可行解 $\bar{\boldsymbol{u}}$，均有

$$f(\bar{\boldsymbol{x}}) = L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}}).$$

推论 13.3 设 $f(\boldsymbol{x})$，$-g_i(\boldsymbol{x})(i=1,2,\cdots,m)$ 均为 $X\subset\mathbf{R}^n$ 上具有一阶连续偏导数的凸函数. 若 $\bar{\boldsymbol{x}}$ 和 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$ 分别为原问题（PP）及对偶问题（DP1）的可行解，且 $f(\bar{\boldsymbol{x}})=L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$，则 $\bar{\boldsymbol{x}}$ 和 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$ 分别为（PP）和（DP1）的最优解.

定理 13.7（对偶定理） 设 $f(\boldsymbol{x})$，$-g_i(\boldsymbol{x})(i=1,2,\cdots,m)$ 均为 $X\subset\mathbf{R}^n$ 上具有一阶连续偏导数的凸函数，$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$ 为 Lagrange 函数 $L(\boldsymbol{x},\boldsymbol{u})$ 的鞍点的充分必要条件为

（1）$\bar{\boldsymbol{x}}$ 是（PP）问题的最优解；

（2）$(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$ 是（DP1）问题的最优解；

（3）$f(\bar{\boldsymbol{x}})=L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$

定理 13.8（强对偶定理） 设 $f(\boldsymbol{x})$，$-g_i(\boldsymbol{x})(i=1,2,\cdots,m)$ 均为 $X\subset\mathbf{R}^n$ 上具有一阶连续偏导数的凸函数. 且 $g_i(\boldsymbol{x})(i=1,2,\cdots,m)$ 满足 Slater 条件，若（PP）存在最优解 $\bar{\boldsymbol{x}}$，则对偶问题（DP1）亦存在最优解 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$，且 $f(\bar{\boldsymbol{x}})=L(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}})$.

第 14 章　可行方向法

可行方向法要求每次迭代都要选择一个下降可行方向及相应的迭代步长，由于下降可行方向有不同的选择，因而对应了不同的可行方向法. 本章将介绍三种可行方向法，即是：

（1）Zoutendijk 可行方向法：通过线性规划来确定下降可行方向；

（2）投影梯度法：把迭代点处的目标函数的负梯度向起作用约束超平面的交集上投影，以确定下降可行方向；

（3）简约梯度法：通过线性等式约束把目标函数变换到低维空间，来确定下降可行方向.

14.1　可行方向法

我们先给出可行方向的定义及几个相关的结论.

定义 14.1　设 $\bar{x}$ 是约束问题

$$\begin{aligned} \min\quad & f(x), x \in \mathbf{R}^n, \\ \text{s.t.}\quad & g_i(x) \geqslant 0, i = 1,2,\cdots,m, \\ & h_i(x) = 0, i = 1,2,\cdots,l \end{aligned} \qquad (14-1)$$

的可行点，可行域为 $D = \{x \mid g_i(x) \geqslant 0, i = 1,2,\cdots,m, h_i(x) = 0, i = 1,2,\cdots,l\}$，若对 $d \neq 0, d \in \mathbf{R}^n$，存在 $\delta > 0$，使当 $\alpha \in (0,\delta]$ 时，有 $\bar{x} + \alpha d \in D$，则称 d 为 $\bar{x}$ 处的一个可行方向.

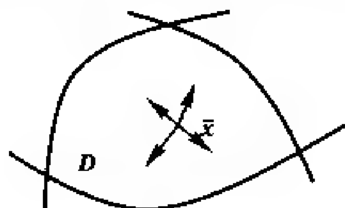


图 14-1

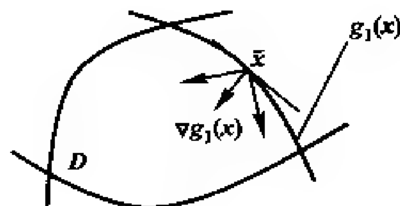


图 14-2

显然，若 $\bar{x}$ 为 D 的内点，则过该点的任意方向都是可行方向，如图 14-1，且约束 $g_i(\bar{x}) > 0$，由定义 12.9 知，此类约束是不起作用的约束；若 $\bar{x}$ 为 D 的边界点，如 $\bar{x}$ 在 D 的边界而 $g_1(x) = 0$ 上，则任何与 $\nabla g_1(\bar{x})$ 夹角小于 90°的方向都是该点处的可行方向，如图 14-2. 而且点 $\bar{x}$ 处于

这个约束形成的可行域的边界上，$\bar{x}$ 的微小变动都可能引起约束的破坏，这一约束起到了限制性的作用，称这种约束为 $\bar{x}$ 处的起作用的约束.

同第 12 章中，我们仍用 $I(\bar{x})$ 表示在可行点 $\bar{x}$ 处起作用的约束的指标集合，即

$$I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0, i = 1,2,\cdots,m\}.$$

例 14.1 考虑由约束

$$\begin{cases} g_1(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 \geqslant 0, \\ g_2(\boldsymbol{x}) = x_1 - x_2 \geqslant 0, \\ g_3(\boldsymbol{x}) = x_1 \geqslant 0, \\ g_4(\boldsymbol{x}) = x_2 \geqslant 0 \end{cases}$$

所确定的可行域（如图 14－3），并讨论点 A,B 处的起作用约束.

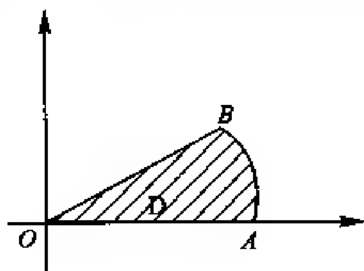


图 14－3

解 通过求解得 A、B 点的坐标为 $\boldsymbol{x}_A = (1,0)^{\mathrm{T}}, \boldsymbol{x}_B = \left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)^{\mathrm{T}}$.

将 $\boldsymbol{x}_A = (1,0)^{\mathrm{T}}$ 代入约束得 $g_1(\boldsymbol{x}_A) = 0, g_4(\boldsymbol{x}_A) = 0$，所以 $g_1(\boldsymbol{x}) \geqslant 0$，$g_4(\boldsymbol{x}) \geqslant 0$ 是关于点 A 的起作用的约束. 即 $I(\boldsymbol{x}_A) = \{1,4\}$. 同理可得 B 点的情形.

下面给出可行方向的几个相关结论.

考虑约束极小问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \boldsymbol{x} \in \mathbf{R}^n, \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \geqslant 0, i = 1,2,\cdots,m, \end{aligned} \tag{14－2}$$

定理 14.1 若 $\boldsymbol{d}$ 为点 $\bar{x} \in \boldsymbol{D}$ 处的可行方向，则有

$$\boldsymbol{d}^{\mathrm{T}} \nabla g_i(\bar{x}) \geqslant 0, i \in I(\bar{x}).$$

证 因为 $\boldsymbol{d}$ 为点 $\bar{x} \in \boldsymbol{D}$ 处的可行方向，所以存在 $\delta > 0$，使当 $\alpha \in$

$(0,\delta]$时,有

$$\bar{x}+\alpha d \in D,$$

即

$$g_i(\bar{x}+\alpha d) \geqslant 0, i=1,2,\cdots,m,$$

而对于 $i \in I(\bar{x})$,仍有

$$g_i(\bar{x}+\alpha d) \geqslant 0,$$

且有

$$g_i(\bar{x})=0,$$

所以,

$$0 \leqslant \frac{d}{d\alpha} g_i(\bar{x}+\alpha d)\bigg|_{\alpha=0} = d^{\mathrm{T}} \nabla g_i(\bar{x}+\alpha d)\bigg|_{\alpha=0}$$
$$= d^{\mathrm{T}} \nabla g_i(\bar{x}), i \in I(\bar{x}).$$

定理 14.2　若 d 满足

$$d^{\mathrm{T}} \nabla g_i(\bar{x}) > 0, i \in I(\bar{x}),$$

则 d 必为点 $\bar{x} \in D$ 处的可行方向.

证　因为

$$g_i(\bar{x})=0, i \in I(\bar{x}),$$

当 $\delta_1>0$ 充分小时,对任意 $\alpha \in (0,\delta_1]$,由 Taylor 展式得

$$\begin{aligned}g_i(\bar{x}+\alpha d) &= g_i(\bar{x})+\alpha d^{\mathrm{T}} \nabla g_i(\bar{x})+o(\alpha)\\ &= \alpha d^{\mathrm{T}} \nabla g_i(\bar{x})+o(\alpha)\\ &= \alpha\left[d^{\mathrm{T}} \nabla g_i(\bar{x})+\frac{o(\alpha)}{\alpha}\right] \geqslant 0, i \in I(\bar{x}),\end{aligned}$$

其中,$o(\alpha)$为比 α 高阶的无穷小量.

当 $i \notin I(\bar{x})$时,因为 $g_i(\bar{x})>0$,由 $g_i(x)$的连续性可知,当 $\delta_2>0$ 充分小时,对任意 $\alpha \in (0,\delta_2]$

亦有

$$g_i(\bar{x}+\alpha d) > 0, i \notin I(x),$$

若取 $\delta=\min(\delta_1,\delta_2)>0$,则对任意 $\alpha \in (0,\delta]$必有

$$g_i(\bar{x}+\alpha d) \geqslant 0, i=1,2,\cdots,m,$$

从而 $\bar{x}+\alpha d \in D$,即 d 为点 $\bar{x} \in D$ 处的可行方向.

已知可行点 $\bar{x} \in D$,若它是极小点,则停止;若它不是极小点,则需要在该点寻找能使目标函数 $f(x)$值下降的可行方向,这样的方向才是我们感兴趣的方向,那么满足什么样的条件才能是下降方向呢?

定义 14.2（下降方向）　若存在向量 $\boldsymbol{d}$ 满足

$$\boldsymbol{d}^{\mathrm{T}}\nabla f(\bar{\boldsymbol{x}})<0,$$

则称方向 $\boldsymbol{d}$ 为 $f(\boldsymbol{x})$ 在点 $\bar{\boldsymbol{x}}\in D$ 处的下降方向.

显然,若 $\boldsymbol{d}$ 为 $f(\boldsymbol{x})$ 在点 $\bar{\boldsymbol{x}}\in D$ 处的下降方向,则必存在 $\delta>0$,使当 $\alpha\in(0,\delta]$ 时,有

$$f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})<f(\bar{\boldsymbol{x}}).$$

总之,在点 $\bar{\boldsymbol{x}}\in D$ 处的可行下降方向应满足

$$\begin{cases}\boldsymbol{d}^{\mathrm{T}}\nabla g_i(\bar{\boldsymbol{x}})\geqslant 0, i\in I(\bar{\boldsymbol{x}}),\\ \boldsymbol{d}^{\mathrm{T}}\nabla f(\bar{\boldsymbol{x}})<0.\end{cases}$$

14.1.1　线性约束的情形

设问题为

$$\begin{aligned}\min\quad & f(\boldsymbol{x}),\boldsymbol{x}\in\mathbf{R}^n\\ \text{s.t.}\quad & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}-b_i=0, i=1,2,\cdots,l,\\ & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}-b_i\geqslant 0, i=1,2,\cdots,m.\end{aligned}\tag{14-3}$$

定理 14.3　设 $\bar{\boldsymbol{x}}$ 是约束问题(14-4)的可行点,则 $\boldsymbol{d}$ 为 $\bar{\boldsymbol{x}}$ 处的可行方向的充分必要条件是:

$$\begin{aligned}&\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}=0, i=1,2,\cdots,l,\\ &\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}\geqslant 0, i\in I(\bar{\boldsymbol{x}}),\end{aligned}\tag{14-4}$$

这里 $I(\bar{\boldsymbol{x}})$ 是 $\bar{\boldsymbol{x}}$ 处的起作用约束指标集.

证　"必要性". 设 $\bar{\boldsymbol{x}}$ 是可行点,$\boldsymbol{d}$ 是 $\bar{\boldsymbol{x}}$ 处的可行方向,则存在 $\delta>0$,使当 $\alpha\in(0,\delta]$ 时,有 $\bar{\boldsymbol{x}}+\alpha\boldsymbol{d}\in\boldsymbol{D}$,即满足

$$\begin{aligned}&\boldsymbol{a}_i^{\mathrm{T}}(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i=0,\quad i=1,2,\cdots,l,\\ &\boldsymbol{a}_i^{\mathrm{T}}(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i\geqslant 0,\quad i=1,2,\cdots,m.\end{aligned}\tag{14-5}$$

因为 $\bar{\boldsymbol{x}}\in\boldsymbol{D}$,当 $i\in\{1,2,\cdots,l\}\cup I(\bar{\boldsymbol{x}})$ 时,有 $\boldsymbol{a}_i^{\mathrm{T}}\bar{\boldsymbol{x}}-b_i=0$,因此(14-4)式成立.

"充分性". 设 $\bar{\boldsymbol{x}}$ 是可行点,$\boldsymbol{d}$ 满足(14-4)式,对 $\forall\alpha$,当 $i=1,2,\cdots,l$ 时,

$$\boldsymbol{a}_i^{\mathrm{T}}(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i=\boldsymbol{a}_i^{\mathrm{T}}\bar{\boldsymbol{x}}-b_i+\alpha\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}=0.\tag{14-6}$$

对 $\forall\alpha\geqslant 0$,当 $i\in I(\bar{\boldsymbol{x}})$ 时

$$\boldsymbol{a}_i^{\mathrm{T}}(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i=\boldsymbol{a}_i^{\mathrm{T}}\bar{\boldsymbol{x}}-b_i+\alpha\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}\geqslant 0.\tag{14-7}$$

当 $i\in\{1,2,\cdots,m\}\setminus I(\bar{\boldsymbol{x}})$ 时,由于只有当 $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}<0$ 时,约束

$$\boldsymbol{a}_i^{\mathrm{T}}(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i=\boldsymbol{a}_i^{\mathrm{T}}\bar{\boldsymbol{x}}-b_i+\alpha\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}\geqslant 0\tag{14-8}$$

才有可能遭到破坏，因此只需取

$$\delta = \min\left\{ -\frac{a_i^T \bar{x} - b_i}{a_i^T d} \middle| a_i^T d < 0, i \notin I(\bar{x}) \right\}.$$

当 $\alpha \in (0,\delta]$ 时，(14－8)式亦成立.

定义 14.3（可行下降方向） 设 $\bar{x}$ 是约束问题(14－3)的可行点，若 d 既是该点处的可行方向，又是下降方向，则称 d 是 $\bar{x}$ 处的可行下降方向.

前面已知，若 d 满足 $\nabla f(\bar{x})^T d < 0$，则 d 是 $\bar{x}$ 处的下降方向. 又由定理 14.3，若要 d 是 $\bar{x}$ 处的可行下降方向，则必须满足：

$$a_i^T d = 0, i = 1,2,\cdots,l, \tag{14-9}$$

$$a_i^T d \geq 0, i \in I(\bar{x}), \tag{14-10}$$

$$\nabla f(\bar{x})^T d < 0. \tag{14-11}$$

因此，我们的目标是寻找满足(14－9)—(14－11)式的 d，得到下降可行方向. 为此，从 $\bar{x}$ 出发，沿方向 d 前进，得到一个新的可行点 $\hat{x}$，而在 $\hat{x}$ 满足

$$f(\hat{x}) < f(\bar{x}),$$

亦即，在满足(14－9)式和(14－10)式的前提下，$\nabla f(\bar{x})^T d$ 越小越好，这就引出一个线性规划问题

$$\begin{aligned} \min \quad & \nabla f(\bar{x})^T d, \\ \text{s.t.} \quad & a_i^T d = 0, i = 1,2,\cdots,l, \\ & a_i^T d \geq 0, i \in I(\bar{x}), \end{aligned} \tag{14-12}$$

则线性规划(14－12)的最优解即是我们所求. 但当 $\alpha \to +\infty$ 时，若有 $\nabla f(\bar{x})^T d \to -\infty$，则求解困难. 为解决这一问题，就必须对 d 或目标函数 $\nabla f(\bar{x})^T d$ 加某些限制（称这些限制为规范约束(Normalization Constraint)），根据规范约束的不同，可以得到不同的规划问题，在这里给出三种规划问题：

问题 Ⅰ

$$\begin{aligned} \min \quad & \nabla f(\bar{x})^T d, \\ \text{s.t.} \quad & a_i^T d = 0, i = 1,2,\cdots,l, \\ & a_i^T d \geq 0, i \in I(\bar{x}), \\ & -1 \leq d_i \leq 1, i = 1,2,\cdots,n. \end{aligned} \tag{14-13}$$

问题 Ⅱ

$$\begin{aligned} \min \quad & \nabla f(\bar{x})^T d, \\ \text{s.t.} \quad & a_i^T d = 0, i = 1,2,\cdots,l, \end{aligned}$$

$$a_i^{\mathrm{T}}\boldsymbol{d} \geqslant 0, i \in I(\bar{x}),$$
$$\boldsymbol{d}^{\mathrm{T}}\boldsymbol{d} \leqslant 1. \tag{14-14}$$

问题Ⅲ

$$\begin{aligned} \min \quad & \nabla f(\bar{x})^{\mathrm{T}}\boldsymbol{d}, \\ \text{s.t.} \quad & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d} = 0, i = 1,2,\cdots,l, \\ & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d} \geqslant 0, i \in I(\bar{x}), \\ & \nabla f(\bar{x})^{\mathrm{T}}\boldsymbol{d} \geqslant -1. \end{aligned} \tag{14-15}$$

在上述三个规划问题中，问题Ⅰ和问题Ⅲ是线性规划问题，因此可用求解线性规划问题的算法求解，问题Ⅱ虽然是一个非线性规划问题，但具有特殊的形式，因此也可用特殊的方法求解.

定理 14.4 设 $\bar{x}$ 是约束问题(14-3)的可行点，则 $\bar{x}$ 成为其 K-T 点的充分必要条件是：问题Ⅰ、Ⅱ或Ⅲ的最优目标函数值为 0.

证 以问题Ⅰ为例，其他两个同理.

充分性 设问题Ⅰ的最优目标函数值为 0，则对满足该问题约束条件的 $\boldsymbol{d}$，均有 $\nabla f(\bar{x})^{\mathrm{T}}\boldsymbol{d} \geqslant 0$，即

$$\begin{aligned} \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d} &= 0, i = 1,2,\cdots,l, \\ \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d} &\geqslant 0, i \in I(\bar{x}), \\ -\nabla f(\bar{x})^{\mathrm{T}}\boldsymbol{d} &> 0 \end{aligned}$$

无解. 由 Farkas 引理，存在 $u_i(i \in I(\bar{x})), v_i(i=1,2,\cdots,l)$，使得

$$\nabla f(\bar{x}) = \sum_{i \in I(\bar{x})} u_i \boldsymbol{a}_i + \sum_{i=1}^{l} v_i \boldsymbol{a}_i,$$
$$u_i \geqslant 0 \quad (i \in I(\bar{x})). \tag{14-16}$$

令 $u_i = 0\ (i \in \{1,2,\cdots,m\} \setminus I(\bar{x}))$，则有

$$\begin{aligned} \nabla f(\bar{x}) - \sum_{i=1}^{m} u_i \boldsymbol{a}_i - \sum_{i=1}^{l} v_i \boldsymbol{a}_i &= 0, \\ u_i \geqslant 0 \quad (i = 1,2,\cdots,m)&, \\ u_i[\boldsymbol{a}_i^{\mathrm{T}}\bar{x} - b_i] > 0, i = 1,2,\cdots,m&, \end{aligned}$$

即 $\bar{x}$ 是 K-T 点.

必要性 设 $\bar{x}$ 是 K-T 点，则有

$$\nabla f(\bar{x}) = \sum_{i=1}^{m} u_i \boldsymbol{a}_i + \sum_{i=1}^{l} v_i \boldsymbol{a}_i.$$

若问题Ⅰ的目标函数值小于 0，则有

$$0 > \nabla f(\bar{x})^{\mathrm{T}}\boldsymbol{d} = \sum_{i=1}^{l} v_i \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d} + \sum_{i \in I(\bar{x})} u_i \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d} + \sum_{i \in \{1,2,\cdots,m\} \setminus I(\bar{x})} u_i \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}$$

$$= \sum_{i \in I(\bar{x})} u_i a_i^T d \geq 0, \tag{14-17}$$

矛盾. 因此问题 I 的最优目标函数值为 0.

由定理 14.4 可知,若问题 I 、II 或 III 的最优目标函数值为 0,则 $\bar{x}$ 是 K－T 点,停止计算;若最优目标函数值小于 0,则 d 是 $\bar{x}$ 处的可行下降方向,那么沿 d 方向进行一维搜索,便得到下一个可行点 $\hat{x}$,并且有 $f(\hat{x}) < f(\bar{x})$,如此下去直到找到最优点.

一维搜索问题

在可行方向法中的一维搜索与无约束问题中的一维搜索是有区别的. 此时的一维搜索除使目标函数值下降外,还要保证其点在可行域内,即在一个区间上求一维极小. 下面给出搜索区间的具体讨论.

设 $x^{(k)}$ 是约束问题(14－3)的可行点,$d^{(k)}$ 是 $x^{(k)}$ 处的可行下降方向,令

$$x = x^{(k)} + \alpha d^{(k)}, \tag{14-18}$$

考虑约束条件,注意 $d^{(k)}$ 是问题 I 、II 或 III 的解,因此有

$$\begin{aligned} a_i^T x - b_i &= a_i^T x^{(k)} - b_i + \alpha a_i^T d^{(k)} = 0, i = 1,2,\cdots,l, \forall \alpha \geq 0, \\ a_i^T x - b_i &= a_i^T x^{(k)} - b_i + \alpha a_i^T d^{(k)} \geq 0, i \in I(x^{(k)}), \forall \alpha \geq 0, \end{aligned} \tag{14-19}$$

因此,可能被破坏的约束是 $x^{(k)}$ 处不起作用的约束

$$a_i^T x - b_i = a_i^T x^{(k)} - b_i + \alpha a_i^T d^{(k)} \geq 0, i \notin I(x^{(k)}), \tag{14-20}$$

其中 $d^{(k)}$ 满足 $a_i^T d^{(k)} < 0$ 的那些约束. 为保证可行性,一维搜索步长应满足

$$\alpha \leq \min\left\{ -\frac{a_i^T x^{(k)} - b_i}{a_i^T d^{(k)}} \middle| a_i^T d^{(k)} < 0, i \notin I(x^{(k)}) \right\},$$

因此取

$$\alpha_{max} = \min\left\{ -\frac{a_i^T x^{(k)} - b_i}{a_i^T d^{(k)}} \middle| a_i^T d^{(k)} < 0, i \notin I(x^{(k)}) \right\}. \tag{14-21}$$

若集合 $\{a_i^T d^{(k)} < 0, i \notin I(x^{(k)})\} = \varnothing$,则令

$$\alpha_{max} = +\infty, \tag{14-22}$$

因此,一维搜索问题为

$$\begin{aligned} \min\ & \phi(\alpha) = f(x^{(k)} + \alpha d^{(k)}), \\ \text{s.t.}\ & 0 \leq \alpha \leq \alpha_{max}. \end{aligned} \tag{14-23}$$

算法 14.1(Zoutendijk 可行方向法)

① 取初始可行点 $x^{(1)}$,即 $x^{(1)}$ 满足

$$a_i^{\mathrm{T}}x^{(1)} - b_i = 0, i = 1,2,\cdots,l,$$
$$a_i^{\mathrm{T}}x^{(1)} - b_i \geqslant 0, \ i = 1,2,\cdots,m,$$

置 $k=1$;

② 确定 $x^{(k)}$ 处起作用约束指标集

$$I(x^{(k)}) = \{i \mid a_i^{\mathrm{T}}x^{(k)} - b_i = 0, \ i = 1,2,\cdots,l\},$$

若 $I(x^{(k)}) = \varnothing$,则 $-\nabla f(x^{(k)}) \Rightarrow d^{(k)}$,转④;否则转③;

③ 求解线性规划子问题

$$\min \quad \nabla f(x^{(k)})^{\mathrm{T}}d,$$
$$\text{s.t.} \ a_i^{\mathrm{T}}d = 0, \ i = 1,2,\cdots,l,$$
$$a_i^{\mathrm{T}}d \geqslant 0, \ i \in I(x^{(k)}),$$
$$-1 \leqslant d_i \leqslant 1, \ i = 1,2,\cdots,n,$$

得到 $d^{(k)}$.

④ 若 $\nabla f(x^{(k)})^{\mathrm{T}}d^{(k)} = 0$,则停止计算,$x^{(k)}$ 为 K-T 点;否则求解一维问题

$$\min \quad \phi(\alpha) = f(x^{(k)} + \alpha d^{(k)})$$
$$\text{s.t.} \quad 0 \leqslant \alpha \leqslant \alpha_{\max},$$

其中

$$\alpha_{\max} = \begin{cases} \min\left\{ -\dfrac{a_i^{\mathrm{T}}x^{(k)} - b_i}{a_i^{\mathrm{T}}d^{(k)}} \,\middle|\, a_i^{\mathrm{T}}d^{(k)} < 0, i \notin I(x^{(k)}), \right\} \\ +\infty, \quad \{a_i^{\mathrm{T}}d^{(k)} < 0, i \notin I(x^{(k)})\} = \varnothing, \end{cases}$$

得到 α_k,置 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$.

⑤ 置 $k = k+1$,转②.

在实际计算中,算法第(4)步 $\nabla f(x^{(k)})^{\mathrm{T}}d^{(k)} = 0$ 可改为 $|\nabla f(x^{(k)})^{\mathrm{T}}d^{(k)}| \leqslant \varepsilon$,其中 ε 是预先指定的精度.

例 14.2 用 Zoutendijk 法求解

$$\min \ (4 - x_2)(x_1 - 3)^2,$$
$$\text{s.t.} \begin{cases} -x_1 - x_2 \geqslant -3, \\ -x_1 \geqslant -2, \\ -x_2 \geqslant -2, \\ x_1 \geqslant 0, \\ x_2 \geqslant 0. \end{cases}$$

解 目标函数的梯度 $\nabla f(x) = [2(4-x_2)(x_1-3), -(x_1-3)^{\mathrm{T}}]$,

令初始可行点 $x^{(0)} = (0.2, 1.8)^{\mathrm{T}}$,

第一次迭代：

（2）确定指标集 $I(\boldsymbol{x}^{(0)})=\varnothing$，$\boldsymbol{x}^{(0)}$ 为可行域的内点，故

$$\boldsymbol{d}^{(0)}=-\nabla f(\boldsymbol{x}^{(0)})=(12.32,7.84)^{\mathrm{T}},$$

（4）$|\nabla f(\boldsymbol{x}^{(0)})^{\mathrm{T}}\boldsymbol{d}^{(0)}|=|-(12.32^2+7.84^2)|>\varepsilon$，则需求解

$$\min\quad \phi(\alpha)=f(\boldsymbol{x}^{(0)}+\alpha\boldsymbol{d}^{(0)}),$$

$$\text{s.t.}\quad 0\leqslant\alpha\leqslant\alpha_{\max},$$

其中，$\alpha_{\max}=\min\left\{\dfrac{1}{20.16},\dfrac{1.8}{12.32},\dfrac{0.2}{7.84}\right\}=\dfrac{0.2}{7.84}=\dfrac{5}{196}$.

$$\min\quad \phi(\alpha)=f(\boldsymbol{x}^{(0)}+\alpha\boldsymbol{d}^{(0)})$$
$$=[4-(1.8+7.8)\alpha][(0.2+12.32\alpha)-3]^2,$$

所以，

$$\text{s.t.}\quad 0\leqslant\alpha\leqslant\frac{5}{196},$$

解之，得
$$\alpha_0=\frac{5}{196}.$$

所以，
$$\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(0)}+\alpha_0\boldsymbol{d}^{(0)}=(0.512,2)^{\mathrm{T}}.$$

第二次迭代：

（2）确定指标集 $I(\boldsymbol{x}^{(1)})=\{3\}$，$\nabla f(\boldsymbol{x}^{(1)})=(-9.95,-6.18)^{\mathrm{T}}$，

（3）求解线性规划子问题

$$\min\ -9.95d_1-6.18d_2,$$

$$\text{s.t.}\quad\begin{cases}-d_2\geqslant 0\\ -1\leqslant d_1\leqslant 1,\\ -1\leqslant d_2\leqslant 1,\end{cases}$$

其最优解为 $\boldsymbol{d}^{(1)}=(1,0)^{\mathrm{T}}$.

（4）$|\nabla f(\boldsymbol{x}^{(1)})^{\mathrm{T}}\boldsymbol{d}^{(1)}|=|-9.95|>\varepsilon$，则需求解

$$\min\quad \phi(\alpha)=f(\boldsymbol{x}^{(1)}+\alpha\boldsymbol{d}^{(1)})=(4-2)(0.512+\alpha-3)^2,$$

$$\text{s.t.}\quad 0\leqslant\alpha\leqslant\alpha_{\max},$$

其中，
$$\alpha_{\max}=\min\{0.488,1.488\}=0.488,$$

解之，得
$$\alpha_1=0.488,$$

所以，
$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\alpha_1\boldsymbol{d}^{(1)}=(1,2)^{\mathrm{T}}.$$

第三次迭代：

（2）确定指标集 $I(\boldsymbol{x}^{(2)})=\{1,3\}$，$\nabla f(\boldsymbol{x}^{(2)})=(-8,-4)^{\mathrm{T}}$.

（3）求解线性规划子问题

$$\min \ -8d_1 - 4d_2,$$

$$\text{s.t.} \begin{cases} -d_1 - d_2 \geqslant 0, \\ -d_2 \geqslant 0, \\ -1 \leqslant d_1 \leqslant 1, \\ -1 \leqslant d_2 \leqslant 1, \end{cases}$$

其最优解为 $d^{(2)} = (1, -1)^T$.

(4) $|\nabla f(x^{(2)})^T d^{(2)}| = |-4| > \varepsilon$,则需求解

$$\min \quad \phi(\alpha) = f(x^{(2)} + \alpha d^{(2)}) = (2+\alpha)(\alpha-2)^2,$$

$$\text{s.t.} \quad 0 \leqslant \alpha \leqslant \alpha_{max}$$

其中，$$\alpha_{max} = \min\{1,2\} = 1,$$

解之,得 $$\alpha_2 = 1,$$

所以，$$x^{(3)} = x^{(2)} + \alpha_2 d^{(2)} = (2,1)^T.$$

第四次迭代：

(2) 确定指标集 $I(x^{(3)}) = \{1,2\}$, $\nabla f(x^{(3)}) = (-6, -1)^T$.

(3) 求解线性规划子问题

$$\min \ -6d_1 - d_2,$$

$$\text{s.t.} \begin{cases} -d_1 - d_2 \geqslant 0, \\ \quad -d_1 \geqslant 0, \\ -1 \leqslant d_1 \leqslant 1, \\ -1 \leqslant d_2 \leqslant 1, \end{cases}$$

其最优解为 $d^{(3)} = (0,0)^T$.

(4) $|\nabla f(x^{(3)})^T d^{(3)}| = 0$,所以,$x^{(3)} = x^{(2)} + \alpha_2 d^{(2)} = (2,1)^T$ 即为 K－T 点.

14.1.2　非线性约束的情形

问题

$$\min \quad f(x), x \in \mathbf{R}^n,$$

$$\text{s.t.} \quad g_i(x) \geqslant 0, \ i = 1,2,\cdots,m, \qquad (14-24)$$

其中,$f(x)$,$g_i(x)$, $i=1,2,\cdots,m$ 是具有一阶连续偏导数的实值函数.

若约束为等式,则只需经过变形可变成两个等式约束,故此处只讨论不等式约束情形.

设经第 k 次迭代已得到可行点 $x^{(k)}$,由前面的讨论,在 $x^{(k)}$ 处当 d 满足如下不等式组时为下降可行方向,即

$$\begin{cases} \boldsymbol{d}^{\mathrm{T}} \nabla g_i(\boldsymbol{x}^{(k)}) > 0, i \in I(\boldsymbol{x}^{(k)}), \\ \boldsymbol{d}^{\mathrm{T}} \nabla f(\boldsymbol{x}^{(k)}) < 0, \end{cases} \tag{14-25}$$

而不等式组(14－25)等价于存在数 σ，使得

$$\begin{cases} -\boldsymbol{d}^{\mathrm{T}} \nabla g_i(\boldsymbol{x}^{(k)}) \leqslant \sigma, i \in I(\boldsymbol{x}^{(k)}), \\ \boldsymbol{d}^{\mathrm{T}} \nabla f(\boldsymbol{x}^{(k)}) \leqslant \sigma, \\ \sigma < 0. \end{cases} \tag{14-26}$$

为了寻求在 $\boldsymbol{x}^{(k)}$ 处的下降可行方向，只需求出满足(14－26)的方向 $\boldsymbol{d}$ 与数 σ. 而对方向 $\boldsymbol{d}$，其本质是要求指出方向，而不要求大小，所以我们限定

$$|d_j| \leqslant 1, j = 1,2,\cdots,n, \boldsymbol{d} = (d_1, d_2, \cdots, d_n)^{\mathrm{T}},$$

容易想到 $|d_j| \leqslant 1, j=1,2,\cdots,n$，且使 $\boldsymbol{d}^{\mathrm{T}} \nabla f(\boldsymbol{x}^{(k)})$ 和 $-\boldsymbol{d}^{\mathrm{T}} \nabla g_i(\boldsymbol{x}^{(k)})$，$i \in I(\boldsymbol{x}^{(k)})$ 的最大值 σ 取极小，从而形成如下形式的线性规划问题.

$$\begin{cases} \min \quad \sigma, \\ \text{s.t.} \quad -\boldsymbol{d}^{\mathrm{T}} \nabla g_i(\boldsymbol{x}^{(k)}) \leqslant \sigma,\ i \in I(\boldsymbol{x}^{(k)}), \\ \qquad\quad \boldsymbol{d}^{\mathrm{T}} \nabla f(\boldsymbol{x}^{(k)}) \leqslant \sigma, \\ \qquad\quad |d_j| \leqslant 1,\ j = 1,2,\cdots,n. \end{cases} \tag{14-27}$$

利用单纯形法求解，设其最优解为 $\begin{bmatrix} \boldsymbol{d}^{(k)} \\ \sigma_k \end{bmatrix}$ 及最优值 σ_k.

而(14－27)所表示的线性规划问题的解，可能是下列两种情形之一：

(1) 若 $\sigma_k = 0$，可以证明不存在方向 $\boldsymbol{d}$ 使 $\boldsymbol{d}^{\mathrm{T}} \nabla g_i(\boldsymbol{x}^{(k)}) > 0$，$i \in I(\boldsymbol{x}^{(k)})$，且 $\boldsymbol{d}^{\mathrm{T}} \nabla f(\boldsymbol{x}^{(k)}) < 0$. 此时，如 $\nabla g_i(\boldsymbol{x}^{(k)})$，$i \in I(\boldsymbol{x}^{(k)})$ 线性无关，则 $\boldsymbol{x}^{(k)}$ 是 K－T 点，迭代停止；

(2) 若 $\sigma_k \neq 0$，由于 $\boldsymbol{d} = 0, \sigma = 0$ 总是它的一个可行解，故在最优解 $\begin{bmatrix} \boldsymbol{d}^{(k)} \\ \sigma_k \end{bmatrix}$ 处，$\sigma_k \leqslant 0$，对 $\sigma_k = 0$. (1) 中已讨论过；对 $\sigma_k < 0$ 有

$$\begin{cases} \boldsymbol{d}^{(k)\mathrm{T}} \nabla f(\boldsymbol{x}^{(k)}) \leqslant \sigma_k < 0, \\ \boldsymbol{d}^{(k)\mathrm{T}} \nabla g_i(\boldsymbol{x}^{(k)}) \geqslant -\sigma_k > 0,\ i \in I(\boldsymbol{x}^{(k)}), \end{cases}$$

故 $\boldsymbol{d}^{(k)}$ 为下降可行方向.

为了确定新的迭代点 $\boldsymbol{x}^{(k+1)}$，从 $\boldsymbol{x}^{(k)}$ 出发沿下降可行方向 $\boldsymbol{d}^{(k)}$ 作受约束的直线搜索，但是，在这里步长因子 α 的上界 $\alpha_{\max}$ 已不像线性约束情形具有显示的计算公式，而是根据新迭代点的可行性由直线搜索来确定，即求解

$$\min_{\alpha\in\Delta} f(x^{(k)}+\alpha d^{(k)}) = f(x^{(k)}+\alpha_k d^{(k)}),$$

$$\Delta = \{\alpha \mid \alpha \geqslant 0,\ x^{(k)}+\alpha_k d^{(k)} \in D\}.$$

设其最优解为 α_k，从而得出新的迭代点 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$. 再重复上面的讨论，进行下一次迭代.

迭代终止条件是：

(1) 若预先给定 σ_k 的允许误差 $\varepsilon_2>0$，由上面讨论的第一种情况知，当 $|\sigma_k| \leqslant \varepsilon_2$ 时，停止迭代；

(2) 若预先给定目标函数的允许误差 $\varepsilon_1>0$，则当

$$\|\nabla f(x^{(k)})\| \leqslant \varepsilon_1$$

时，停止迭代.

最后还需指出，若 $I(x^{(k)})=\varnothing$，则 $-\nabla f(x^{(k)}) \Rightarrow d^{(k)}$，即取最速下降方向作为下降可行方向.

归纳以上的讨论，可得出一般约束问题的可行方向算法.

算法 14.2(Zoutendijk 算法)

① 选定初始可行点 $x^{(0)}$，给定目标函数允许误差限 $\varepsilon_1>0$ 及 σ_k 的允许误差 $\varepsilon_2>0$，$0\Rightarrow k$；

② 确定指标集

$$I(x^{(k)}) = \{i \mid g_i(x^{(k)}) = 0,\ 1\leqslant i\leqslant m\};$$

③ 检验条件 $I(x^{(k)})=\varnothing$，若满足，且 $\|\nabla f(x^{(k)})\| \leqslant \varepsilon_1$，停止迭代，则 $x^{(k)}$ 为极小点，若 $\|\nabla f(x^{(k)})\| > \varepsilon_1$，则令 $-\nabla f(x^{(k)}) \Rightarrow d^{(k)}$，转⑥；若 $I(x^{(k)}) \neq \varnothing$，则转④；

④ 求解线性规划问题

$$\min\ \sigma,$$

$$\begin{cases} d^{\mathrm{T}}\nabla f(x^{(k)}) \leqslant \sigma, \\ -d^{\mathrm{T}}\nabla g_i(x^{(k)}) \leqslant \sigma,\ i\in I(x^{(k)}), \\ |d_j| \leqslant 1, j=1,2,\cdots,n, \end{cases}$$

得到 $d^{(k)}$，σ_k；

⑤ 检验终止准则

$$|\sigma_k| \leqslant \varepsilon_2,$$

若满足，则迭代终止，$x^{(k)}$ 为极小点，否则，转⑥；

⑥ 求解一维问题

$$\min_{\alpha\in\Delta} = f(x^{(k)}+\alpha d^{(k)}),$$

其中　　$$\Delta = \{\alpha \mid \alpha\geqslant 0,\ x^{(k)}+\alpha d^{(k)} \in D\},$$

解之，得 α_k；

⑦ 置 $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}$，$k=k+1$，转②.

计算实践和理论分析表明算法 14.2 可能会导致失效或出现“拉锯现象”，致使可行方向法收敛很慢乃至不收敛到最优点或 K－T 点. 因此研究如何有效的防止“拉锯现象”的出现便成为十分必要的.

算法在形成找可行方向的问题时，对于线性和非线性不等式约束两者都仅仅只考虑了 $x^{(k)}$ 处起作用的约束，这时求出来的可行方向有可能很快就碰上靠近 $x^{(k)}$ 的某个约束边界. 如 $g_2(x)=0$，由于这边界很靠近 $x^{(k)}$，当 $d^{(k)}$ 取得不恰当时，沿 $d^{(k)}$ 进行一维搜索只可能移动一个很小的步长，否则就会使迭代点跑出边界. 为了防止这些现象的出现，采用 ε－起作用可行方向法，这种方法在形成找可行方向问题时不仅把通过当前点 $x^{(k)}$ 的那些约束边界看作是起作用的约束，而且把充分靠近 $x^{(k)}$ 的那些约束也看作是起作用约束，即把满足 $0\leqslant g_i(x^{(k)})\leqslant\varepsilon_k$ 的约束也考虑在内，这样的约束称为 ε－起作用约束，而对应的可行方向法称为 ε－起作用约束可行方向法.

算法 14.3（ε－起作用约束可行方向法）

① 选定初始可行点 $x^{(0)}$，给定 $\varepsilon_0>0$ 及 $\varepsilon>0$，$0\Rightarrow k$；

② 确定 ε_k 起作用约束指标集

$$I(x^{(k)},\varepsilon_k)=\{i\mid 0\leqslant g_i(x^{(k)})\leqslant\varepsilon_k,1\leqslant i\leqslant m\},$$

计算 $\nabla f(x^{(k)})$；

③ 检验条件 $I(x^{(k)},\varepsilon_k)=\varnothing$，若满足，且 $\|\nabla f(x^{(k)})\|\leqslant\varepsilon$，停止迭代，则 $x^{(k)}$ 为近似极小点，若 $\|\nabla f(x^{(k)})\|>\varepsilon$，则令 $-\nabla f(x^{(k)})\Rightarrow d^{(k)}$，$\sigma_k=-1$ 转⑥；若 $I(x^{(k)},\varepsilon_k)\neq\varnothing$，则转④；

④ 求解线性规划问题

$$\min\quad\sigma,$$
$$\begin{cases}d^{\mathrm T}\nabla f(x^{(k)})\leqslant\sigma,\\-d^{\mathrm T}\nabla g_i(x^{(k)})\leqslant\eta_i\sigma,\ i\in I(x^{(k)},\varepsilon_k),\\|d_j|\leqslant 1,\ j=1,2,\cdots,n.\end{cases}$$

当 ε－起作用约束为线性约束时，取 $\eta_i=0$；当 ε－起作用约束为非线性约束时，取 $\eta_i=1$，解之得到 $d^{(k)}$，σ_k；

⑤ 检验终止准则，若 $\sigma_k=0$，且 $\varepsilon_k<\varepsilon$，则 $x^{(k)}$ 为近似极小点，迭代终止，否则，$\varepsilon_k\geqslant\varepsilon$，则 $\dfrac{\varepsilon_k}{2}\Rightarrow\varepsilon_k$ 转②；

⑥ 以 $d^{(k)}$ 为方向进行一维搜索. 先对 $i\notin I(x^{(k)},\varepsilon_k)$，求 $t_i>0$，使

$g_i(\boldsymbol{x}^{(k)}+t_i\boldsymbol{d}^{(k)})=0$（若 $t_i>0$ 不止一个，则取最小的一个）即

$$\alpha_{\max}=\min\{t_i \mid g_i(\boldsymbol{x}^{(k)}+t_i\boldsymbol{d}^{(k)})=0, t_i>0, i\notin I(\boldsymbol{x}^{(k)},\varepsilon_k)\},$$

同时求解

$$\begin{cases}\min f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)}),\\ 0\leqslant\alpha\leqslant\alpha_{\max},\end{cases}$$

解之，得 α_k；置 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{d}^{(k)}$；

⑦ 若 $\|\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}\|<\varepsilon$，则 $\boldsymbol{x}^{(k+1)}$ 为近似最优解，迭代终止.

否则，令 $\varepsilon_k=\begin{cases}\varepsilon_k, \varepsilon_k\leqslant-\sigma_k,\\ \dfrac{\varepsilon_k}{2}, \varepsilon_k>-\sigma_k,\end{cases}$ $k+1\Rightarrow k$，转②.

例 14.3 用 ε - 起作用约束可行方向法求解

$$\min f(\boldsymbol{x})=(x_1-3)^2(4-x_2),$$

$$\text{s.t.}\begin{cases}g_1(\boldsymbol{x})=4-x_1^2-x_2^2+2x_1\geqslant0,\\ g_2(\boldsymbol{x})=5-x_1^2-x_2^2\geqslant0,\\ g_3(\boldsymbol{x})=4-x_1^2-x_2^2+2x_2\geqslant0,\\ g_4(\boldsymbol{x})=x_1\geqslant0,\\ g_5(\boldsymbol{x})=x_2\geqslant0.\end{cases}$$

解 目标函数的梯度 $\nabla f(\boldsymbol{x})=[2(4-x_2)(x_1-3),-(x_1-3)^2]$，

令初始可行点 $\boldsymbol{x}^{(0)}=\left(\dfrac{1}{2},1\right)^{\mathrm{T}}$，$\varepsilon_0=\dfrac{1}{4}$，$\varepsilon=0.000\ 1$，

第一次迭代：

（2）确定指标集 $I(\boldsymbol{x}^{(0)},\varepsilon_0)=\varnothing$，故

$$\nabla f(\boldsymbol{x}^{(0)})=(-15,-6.25)^{\mathrm{T}};$$

（6）沿 $\boldsymbol{d}^{(0)}=-\nabla f(\boldsymbol{x}^{(0)})=(15,6.25)^{\mathrm{T}}$ 方向进行一维搜索，先求 $t_i>0$，使

$$g_i(\boldsymbol{x}^{(0)}+t_i\boldsymbol{d}^{(0)})=0,\ i=1,2,3,$$

由 $g_1(\boldsymbol{x}^{(0)}+t_1\boldsymbol{d}^{(0)})=4-\left(\dfrac{1}{2}+15t_1\right)^2-(1+6.25t_1)^2+2\left(\dfrac{1}{2}+15t_1\right)=0,$

求得 $t_1=0.124$，同理得 $t_2=0.077\ 97$，$t_3=0.165\ 5$，

所以，$$\alpha_{\max}=\min\{t_1,t_2,t_3\}=0.077\ 97,$$

故 $$\begin{cases}\min\quad f(\boldsymbol{x}^{(0)}+\alpha\boldsymbol{d}^{(0)})=\left(15\alpha-\dfrac{5}{2}\right)^2(3-6.25\alpha),\\ \text{s.t.}\quad 0\leqslant\alpha\leqslant\alpha_{\max},\end{cases}$$

解之，得 $$\alpha_0=0.077\ 97,$$

所以，$$x^{(1)}=x^{(0)}+\alpha_0 d^{(0)}=(1.67,1.487)^T,$$

第二次迭代

$$\varepsilon_0\Rightarrow\varepsilon_1=0.25.$$

(2) 确定指标集 $I(x^{(1)},\varepsilon_1)=\{2\}\neq\varnothing$,

(4) 求解线性规划问题

$$\begin{cases}\min\quad \sigma,\\ \text{s.t.}\quad d^T\nabla g_2(x^{(1)})=-3.34d_1-2.974d_2\geqslant-\sigma,\\ \qquad d^T\nabla f(x^{(1)})=-6.68458d_1-1.7689d_2\leqslant\sigma,\\ \qquad |d_1|\leqslant1,\\ \qquad |d_2|\leqslant1.\end{cases}$$

解得 $d^{(1)}=(0.473,-1)^T,\sigma_1=-1.39<0$,所以,$d^{(1)}$为所求可行方向;

(5) 检验终止准则,$\sigma_1=-1.39<0$;

(6) 以 $d^{(1)}$为方向进行一维搜索

$$x^{(1)}+\alpha d^{(1)}=(1.67+0.473\alpha,1.487-\alpha)^T,$$

$$\alpha_{max}=\min\{t_i\mid g_i(x^{(k)}+t_id^{(k)})=0,\ t_i>0,\ i=1,3,5\}=t_2=1.0464,$$

故 $$\begin{cases}\min\quad f(x^{(1)}+\alpha d^{(1)})=(0.473\alpha-1.33)^2(2.513+\alpha),\\ \text{s.t.}\quad 0\leqslant\alpha\leqslant\alpha_{max},\end{cases}$$

解之,得 $$\alpha_1=1.0464,$$

所以，$$x^{(2)}=x^{(1)}+\alpha_1 d^{(1)}=(2.1649,0.441)^T.$$

(7) $\varepsilon_1=0.25<-\sigma=1.39$,所以,令 $\varepsilon_1\Rightarrow\varepsilon_2=0.25$,

第三次迭代:

(2) 确定指标集 $I(x^{(2)},\varepsilon_2)=\{2,3\}\neq\varnothing$;

(4) 求解线性规划问题

$$\begin{cases}\min\quad \sigma,\\ \text{s.t.}\quad d^T\nabla g_2(x^{(2)})=-4.33d_1-0.88d_2\geqslant-\sigma,\\ \qquad d^T\nabla g_3(x^{(2)})=-4.33d_1+1.12d_2\geqslant-\sigma,\\ \qquad d^T\nabla f(x^{(2)})=-5.9d_1-0.7d_2\leqslant\sigma,\\ |d_1|\leqslant1,\\ |d_2|\leqslant1,\end{cases}$$

解得 $$d^{(2)}=(0,0)^T,\sigma_2=0,$$

(5) 检验终止准则,$\varepsilon_2>\varepsilon$,所以,$\frac{\varepsilon_2}{2}=0.125\Rightarrow\varepsilon_2$,转(2);

(2) 确定指标集 $I(x^{(2)},\varepsilon_2)=\{2,3\}\neq\varnothing$,

(4) 求解线性规划问题

$$\begin{cases} \min \quad \sigma, \\ \text{s.t.} \quad d^T \nabla g_2(x^{(2)}) = -4.33d_1 - 0.88d_2 \geqslant -\sigma, \\ \quad d^T \nabla g_3(x^{(2)}) = -4.33d_1 + 1.12d_2 \geqslant -\sigma, \\ \quad d^T \nabla f(x^{(2)}) = -5.9d_1 - 0.7d_2 \leqslant \sigma, \\ \quad |d_1| \leqslant 1, \\ \quad |d_2| \leqslant 1, \end{cases}$$

解得　　　　$d^{(2)} = (0,0)^T, \sigma_2 = 0$,

(5) 检验终止准则,$\varepsilon_2 > \varepsilon$,所以,$\frac{\varepsilon_2}{2} = \frac{1}{16} \Rightarrow \varepsilon_2$,转(2);

(2) 确定指标集 $I(x^{(2)}, \varepsilon_2) = \{3\} \neq \varnothing$,

(4) 求解线性规划问题

$$\begin{cases} \min \quad \sigma, \\ \text{s.t.} \quad d^T \nabla g_3(x^{(2)}) = -4.33d_1 + 1.12d_2 \geqslant -\sigma, \\ \quad d^T \nabla f(x^{(2)}) = -5.9d_1 - 0.7d_2 \leqslant \sigma, \\ \quad |d_1| \leqslant 1, \\ \quad |d_2| \leqslant 1, \end{cases}$$

解得 $d^{(2)} = (0.041, 1)^T, \sigma_2 = -0.924 < 0$,所以,$d^{(2)}$ 为所求可行方向;

(5) 检验终止准则,$\sigma_2 = -0.924 < 0$;

(6) 以 $d^{(2)}$ 为方向进行一维搜索

$$x^{(2)} + \alpha d^{(2)} = (2.1649 + 0.041\alpha, 0.041 + \alpha)^T,$$

$$\alpha_{max} = \min\{t_i \mid g_i(x^{(k)} + t_i d^{(k)}) = 0, t_i > 0, i = 1,2\} = t_2 = 0.103,$$

故　　$\min f(x^{(2)} + \alpha d^{(2)}) = (0.041\alpha - 0.8351)^2(3.559 - \alpha)$,

$\text{s.t.} \quad 0 \leqslant \alpha \leqslant \alpha_{max}$,

解之,得　　　　$\alpha_2 = 0.103$,

所以,　　　　$x^{(3)} = x^{(2)} + \alpha_2 d^{(2)} = (2.169, 0.544)^T$.

(7) $\varepsilon_2 = \frac{1}{16} < -\sigma = 0.924$,所以,令 $\varepsilon_2 \Rightarrow \varepsilon_3 = \frac{1}{16}$(最优解为:$g_2(x) = 0, g_3(x) = 0$ 之交点 $x^* = \left(\frac{\sqrt{19}}{2}, \frac{1}{2}\right)^T$).

若继续迭代下去,可看出点列 $\{x^{(k)}\}$ 是逐步逼近 x^* 的,但收敛的速度是很慢的. 这也说明可行方向法对于非线性约束较多的问题的实现是有困难的.

14.2　投影梯度法

投影梯度法的基本思想是，将负梯度方向投影到可行方向上去.

定义 14.4　若 $n\times n$ 阶矩阵 $\boldsymbol{P}$ 满足

$$\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{P}\ \&\ \boldsymbol{P}^2=\boldsymbol{P}, \tag{14-28}$$

则称 $\boldsymbol{P}$ 为投影矩阵.

定理 14.5　设 $\boldsymbol{P}$ 是 $n\times n$ 阶矩阵，则有下述结论：

(1) 若 $\boldsymbol{P}$ 是投影矩阵，则 $\boldsymbol{P}$ 为半正定矩阵；

(2) $\boldsymbol{P}$ 是投影矩阵的充分必要条件是：$\boldsymbol{I}-\boldsymbol{P}$ 为投影矩阵；

(3) 设 $\boldsymbol{P}$ 是投影矩阵，令 $\boldsymbol{Q}=\boldsymbol{I}-\boldsymbol{P}$，记

$$L=\{\boldsymbol{Px}\mid\boldsymbol{x}\in\mathbf{R}^n\}\ 和\ L^{\perp}=\{\boldsymbol{Qx}\mid\boldsymbol{x}\in\mathbf{R}^n\}, \tag{14-29}$$

则 L 和 $L^{\perp}$ 是正交的线性空间，且任一点 $\boldsymbol{x}\in\mathbf{R}^n$，可唯一地表示成 $\boldsymbol{P}+\boldsymbol{Q}$，其中 $\boldsymbol{P}\in L,\boldsymbol{Q}\in L^{\perp}$.

定理 14.6　设 $\bar{\boldsymbol{x}}$ 是约束问题(14-3)的可行点，且 $f(\boldsymbol{x})$ 又有连续的一阶偏导数，若 $\boldsymbol{P}$ 是投影矩阵，且 $\boldsymbol{P}\nabla f(\bar{\boldsymbol{x}})\neq 0$，则

$$\boldsymbol{d}=-\boldsymbol{P}\nabla f(\bar{\boldsymbol{x}}), \tag{14-30}$$

此外，若

$$\boldsymbol{N}=(\boldsymbol{a}_{i_1},\boldsymbol{a}_{i_2},\cdots,\boldsymbol{a}_{i_r}) \tag{14-31}$$

为列满秩矩阵. 其中 $i_j\in\{1,2,\cdots,l\}\cup I(\bar{\boldsymbol{x}})\ (j=1,2,\cdots,5)$，令 $\boldsymbol{P}$ 具有如下形式：

$$\boldsymbol{P}=\boldsymbol{I}-\boldsymbol{N}(\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N})^{-1}\boldsymbol{N}^{\mathrm{T}}, \tag{14-32}$$

则 $\boldsymbol{d}$ 是 $\bar{\boldsymbol{x}}$ 处的可行下降方向.

证　因为 $\boldsymbol{P}\nabla f(\bar{\boldsymbol{x}})\neq 0$，则有

$$\begin{aligned}\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d}&=-\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{P}\nabla f(\bar{\boldsymbol{x}})\\&=-\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\nabla f(\bar{\boldsymbol{x}})\\&=-\|\boldsymbol{P}\nabla f(\bar{\boldsymbol{x}})\|^2,\end{aligned} \tag{14-33}$$

因此，$\boldsymbol{d}$ 是下降方向.

若 $\boldsymbol{P}$ 具有(14-32)式的形式，则有

$$\boldsymbol{N}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{N}^{\mathrm{T}}-\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N}(\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N})^{-1}\boldsymbol{N}^{\mathrm{T}}=0, \tag{14-34}$$

因此有

$$\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}=-\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{P}\nabla f(\bar{\boldsymbol{x}})=0,\ i=\{1,2,\cdots,l\}\cup I(\bar{\boldsymbol{x}}), \tag{14-35}$$

则 $\boldsymbol{d}$ 是可行方向，因此 $\boldsymbol{d}$ 是可行下降方向.

由定理 14.6 可知，在 $P\nabla f(\bar{x})\neq 0$ 的条件下，可以得到 $\bar{x}$ 处的可行下降方向，进而便可得到下一个可行点. 当 $P\nabla f(\bar{x})=0$ 时，以下定理说明之.

定理 14.7 设 $\bar{x}$ 是约束问题(14-3)的可行点，且 $f(x)$ 有连续的一阶偏导数，若 P 是由(14-31)式和(14-32)式确定的矩阵，并满足

$$P\nabla f(\bar{x})=0, \tag{14-36}$$

记

$$\lambda=(N^{T}N)^{-1}N^{T}\nabla f(\bar{x}), \tag{14-37}$$

若 $\lambda_i\geqslant 0, i\in I(\bar{x})$，则 $\bar{x}$ 是 K-T 点. 若存在 $q\in I(\bar{x})$，使得 $\lambda_q<0$，并记 $\bar{N}$ 为矩阵 N 中去掉 λ_q 对应的列后得到的矩阵，并令 $\bar{P}=I-\bar{N}(\bar{N}^{T}\bar{N})^{-1}\bar{N}^{T}$，$d=-\bar{P}\nabla f(\bar{x})$，则 d 是 $\bar{x}$ 处的可行下降方向.

证 由(14-32)式、(14-36)式和(14-37)式，有

$$\begin{aligned}0=P\nabla f(\bar{x})&=(I-N(N^{T}N)^{-1}N^{T})\nabla f(\bar{x})\\&=\nabla f(\bar{x})-N(N^{T}N)^{-1}N^{T}\nabla f(\bar{x})\\&=\nabla f(\bar{x})-N\lambda,\end{aligned} \tag{14-38}$$

即

$$\nabla f(\bar{x})-\sum_{i\in\{1,2,\cdots,l\}\cup I(\bar{x})}\lambda_i a_i=0. \tag{14-39}$$

若 $\lambda_i\geqslant 0, i\in I(\bar{x})$，则令 $\lambda_i=0, i\notin I(\bar{x})$，(14-39)式即为 K-T 条件，所以 $\bar{x}$ 是 K-T 点.

若存在 $q\in I(\bar{x})$，使得 $\lambda_q<0$，证明 $d=-\bar{P}\nabla f(\bar{x})$ 是可行下降方向.

(1) 证明 $d\neq 0$，(反证法)若 $d=0$，则由 $P\nabla f(\bar{x})=0$，必存在

$$\mu=(\bar{N}^{T}\bar{N})^{-1}N^{T}\nabla f(\bar{x}),$$

使得

$$\nabla f(\bar{x})-\sum_{\substack{i\in\{1,2,\cdots,l\}\cup I(\bar{x})\\ i\neq q}}\mu_i a_i=0 \tag{14-40}$$

用(14-39)式减去(14-40)式，得到

$$\lambda_q a_q-\sum_{\substack{i\in\{1,2,\cdots,l\}\cup I(\bar{x})\\ i\neq q}}(\lambda_i-\mu_i)a_i=0, \tag{14-41}$$

由于 $\lambda_q<0$，这与 N 为列满秩矩阵矛盾.

(2) 证明 d 是可行方向：类似于定理(14-29)的证明可知

$$a_i^{T}d=0,\ i\in\{1,2,\cdots,l\}\cup I(\bar{x}), i\neq q, \tag{14-42}$$

并由(14-39)式和(14-42)式，有

$$a_q^{T}d=\left(\frac{1}{\lambda_q}\nabla f(\bar{x})-\frac{1}{\lambda_q}\sum_{\substack{i\in\{1,2,\cdots,l\}\cup I(\bar{x})\\ i\neq q}}\lambda_i a_i\right)^{T}d$$

$$
\begin{aligned}
&= \frac{1}{\lambda_q} \nabla f(\bar{x})^{\mathrm{T}} d \\
&= \frac{1}{\lambda_q} \nabla f(\bar{x})^{\mathrm{T}} \overline{P} \nabla f(\bar{x}) \qquad (14-43) \\
&= \frac{1}{\lambda_q} \| \overline{P} \nabla f(\bar{x}) \|^2,
\end{aligned}
$$

所以 d 是 $\bar{x}$ 处的可行方向. 由定理 14.6 可知, $d \neq 0$ 是下降方向.

由定理 14.6、14.7 可以给出投影梯度法的具体算法.(由于投影梯度法是 J. B. Rosen 提出来的, 因此也称为 Rosen 投影梯度法.)

算法 14.4(Rosen 投影梯度法)

① 取初始可行点 $x^{(1)}$, 即 $x^{(1)}$ 满足

$$
a_i^{\mathrm{T}} x^{(1)} - b_i = 0,\ i = 1,2,\cdots,l,
$$
$$
a_i^{\mathrm{T}} x^{(1)} - b_i \geqslant 0,\ i = 1,2,\cdots,m,
$$

置 $k=1$;

② 确定 $x^{(k)}$ 处的起作用约束指标集

$$
I(x^{(k)}) = \{ i \mid a_i^{\mathrm{T}} x^{(k)} - b_i = 0,\ i = 1,2,\cdots,m \},
$$

③ 若 $l=0$(无等式约束)且 $I(x^{(k)}) = \varnothing$, 则令

$$
d^{(k)} = - \nabla f(x^{(k)}),
$$

否则令

$$
N^{(k)} = (a_{i_1}, a_{i_2}, \cdots, a_{i_r}),\ i_j \in \{1,2,\cdots,l\} \cup I(x^{(k)}),
$$
$$
P^{(k)} = I - N^{(k)} (N^{(k)\mathrm{T}} N^{(k)})^{-1} N^{(k)\mathrm{T}},
$$
$$
d^{(k)} = - P^{(k)} \nabla f(x^{(k)}).
$$

④ 若 $d^{(k)} = 0$(一般为: $\| d^{(k)} \| \leqslant \varepsilon$),

(a) 若 $l=0$, 且 $I(x^{(k)}) = \varnothing$, 则停止计算, $x^{(k)}$ 为 K-T 点; 否则计算

$$
\lambda^{(k)} = (N^{(k)\mathrm{T}} N^{(k)})^{-1} N^{(k)\mathrm{T}} \nabla f(x^{(k)}),
$$

(b) $\lambda_i^{(k)} \geqslant 0, i \in I(x^{(k)})$, 则停止计算, $x^{(k)}$ 为 K-T 点; 否则令

$$
\lambda_q^{(k)} = \min \{ \lambda_i^{(k)} \mid i \in I(x^{(k)}) \},
$$

置

$$
N^{(k)} = (N^{(k)} \text{ 中去掉 } \lambda_q^{(k)} \text{ 对应的列 } a_q),
$$
$$
P^{(k)} = I - N^{(k)} (N^{(k)\mathrm{T}} N^{(k)})^{-1} N^{(k)\mathrm{T}},
$$
$$
d^{(k)} = - P^{(k)} \nabla f(x^{(k)}).
$$

⑤ $(d^{(k)} \neq 0)$ 求解一维问题

$$
\min \varphi(\alpha) = f(x^{(k)} + \alpha d^{(k)}),
$$
$$
\text{s.t.} \quad 0 \leqslant \alpha \leqslant \alpha_{\max},
$$

其中

$$\alpha_{\max}=\begin{cases}\min\left\{\dfrac{b_i-a_i^{\mathrm T}x^{(k)}}{a_i^{\mathrm T}d^{(k)}}\ \middle|\ a_i^{\mathrm T}d^{(k)}<0,\ i\notin I(x^{(k)})\right\}\\ +\infty\qquad,\{a_i^{\mathrm T}d^{(k)}<0,i\notin I(x^{(k)})\}=\varnothing,\end{cases}$$

得到 α_k,置 $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}$,

⑥ 置 $k=k+1$,转②.

例 14.4 用投影梯度法求解

$$\begin{cases}\min & x_1^2+4x_2^2,\\ \text{s.t.} & x_1+x_2\geqslant 1,\\ & 15x_1+10x_2\geqslant 12,\\ & x_1\geqslant 0,\\ & x_2\geqslant 0.\end{cases}$$

解 (1) 取初始可行点 $x^{(1)}=\begin{bmatrix}0\\2\end{bmatrix}$,即 $x^{(1)}$ 满足

$$a_i^{\mathrm T}x^{(1)}-b_i=0,\ i=1,2,\cdots,l,$$
$$a_i^{\mathrm T}x^{(1)}-b_i\geqslant 0,\ i=1,2,\cdots,m,$$

置 $k=1$;

(2) 确定 $x^{(1)}$ 处的起作用约束指标集

$$I(x^{(1)})=\{3\},$$

(3) $l=0$(无等式约束),但 $I(x^{(1)})\neq\varnothing$,则令

$$N^{(1)}=\begin{bmatrix}1\\0\end{bmatrix},$$

$$\begin{aligned}P^{(1)}&=I-N^{(1)}(N^{(1)\mathrm T}N^{(1)})^{-1}N^{(1)\mathrm T}\\&=\begin{bmatrix}1&0\\0&1\end{bmatrix}-\begin{bmatrix}1\\0\end{bmatrix}\left(\begin{bmatrix}1&0\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\right)^{-1}\begin{bmatrix}1&0\end{bmatrix}\\&=\begin{bmatrix}0&0\\0&1\end{bmatrix},\end{aligned}$$

$$\begin{aligned}d^{(1)}&=-P^{(1)}\ \nabla f(x^{(1)})\\&=-\begin{bmatrix}0&0\\0&1\end{bmatrix}\begin{bmatrix}0\\16\end{bmatrix}\\&=\begin{bmatrix}0\\-16\end{bmatrix}\neq 0,\end{aligned}$$

(4) $d^{(1)}=\begin{bmatrix}0\\-16\end{bmatrix}\neq\mathbf{0}$,所以,求解一维问题

$$\min \quad \phi(\alpha) = f(\boldsymbol{x}^{(1)} + \alpha \boldsymbol{d}^{(1)}),$$
$$\text{s.t.} \quad 0 \leqslant \alpha \leqslant \alpha_{\max},$$

其中

$$\alpha_{\max} = \min\left\{\frac{1-2}{-16}, \frac{12-20}{-160}, \frac{0-2}{-16}\right\} = \frac{1}{20},$$

即

$$\min_{\alpha \in [0, \frac{1}{20}]} (4(2-16\alpha)^2),$$

解得 $\alpha_1 = \frac{1}{20}$，置 $\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \alpha_1 \boldsymbol{d}^{(1)} = \begin{bmatrix} 0 \\ \frac{6}{5} \end{bmatrix}$.

第二次迭代：

（2）确定 $\boldsymbol{x}^{(2)}$ 处的起作用约束指标集

$$I(\boldsymbol{x}^{(2)}) = \{2,3\},$$

（3）$l=0$（无等式约束），但 $I(\boldsymbol{x}^{(2)}) \neq \varnothing$，则令

$$\boldsymbol{N}^{(2)} = \begin{bmatrix} 15 & 1 \\ 10 & 0 \end{bmatrix},$$

$$\begin{aligned}
\boldsymbol{P}^{(2)} &= \boldsymbol{I} - \boldsymbol{N}^{(2)}(\boldsymbol{N}^{(2)\mathrm{T}}\boldsymbol{N}^{(2)})^{-1}\boldsymbol{N}^{(2)\mathrm{T}} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 15 & 1 \\ 10 & 0 \end{bmatrix}\left(\begin{bmatrix} 15 & 10 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 15 & 1 \\ 10 & 0 \end{bmatrix}\right)^{-1}\begin{bmatrix} 15 & 10 \\ 1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{d}^{(2)} &= -\boldsymbol{P}^{(2)} \nabla f(\boldsymbol{x}^{(2)}) \\
&= -\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 48/5 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\end{aligned}$$

（4）因为 $\boldsymbol{d}^{(2)} = \boldsymbol{0}$，

（a）$l=0$，但 $I(\boldsymbol{x}^{(2)}) \neq \varnothing$ 则

$$\begin{aligned}
\boldsymbol{\lambda}^{(2)} &= (\boldsymbol{N}^{(2)\mathrm{T}}\boldsymbol{N}^{(2)})^{-1}\boldsymbol{N}^{(2)\mathrm{T}} \nabla f(\boldsymbol{x}^{(2)}) \\
&= \begin{bmatrix} 0.01 & -0.15 \\ -0.15 & 3.25 \end{bmatrix}\begin{bmatrix} 15 & 10 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 48/5 \end{bmatrix} \\
&= \begin{bmatrix} 0.96 \\ -14.4 \end{bmatrix},
\end{aligned}$$

（b）$\lambda_q^{(2)} = \min\{\lambda_i^{(2)} \mid i \in I(\boldsymbol{x}^{(2)})\} = -14.4$，

置

$$N^{(2)}=\begin{bmatrix}15\\10\end{bmatrix},$$

$$\begin{aligned}P^{(2)}&=I-N^{(2)}(N^{(2)T}N^{(2)})^{-1}N^{(2)T}\\&=\begin{bmatrix}1&0\\0&1\end{bmatrix}-\begin{bmatrix}15\\10\end{bmatrix}\left(\begin{bmatrix}15&10\end{bmatrix}\begin{bmatrix}15\\10\end{bmatrix}\right)^{-1}\begin{bmatrix}15&10\end{bmatrix}\\&=\begin{bmatrix}\dfrac{100}{325}&-\dfrac{150}{325}\\-\dfrac{150}{325}&-\dfrac{225}{325}\end{bmatrix},\end{aligned}$$

$$\begin{aligned}d^{(2)}&=-P^{(2)}\nabla f(x^{(2)})\\&=\begin{bmatrix}\dfrac{100}{325}&-\dfrac{150}{325}\\-\dfrac{150}{325}&-\dfrac{225}{325}\end{bmatrix}\begin{bmatrix}0\\\dfrac{48}{5}\end{bmatrix}\\&=\begin{bmatrix}4.4304\\-6.6461\end{bmatrix}.\end{aligned}$$

(5) $d^{(2)}\neq 0$,所以,求解一维问题

$$\begin{aligned}\min\quad&\varphi(\alpha)=f(x^{(2)}+\alpha d^{(2)}),\\\text{s.t.}\quad&0\leqslant\alpha\leqslant\alpha_{\max},\end{aligned}$$

其中

$$\alpha_{\max}=\min\left\{\frac{1-(6/5)}{-2.2157},\frac{-6/5}{-6.6461}\right\}=0.09,$$

即
$$\min_{\alpha\in[0,0.09]}(4.4304\alpha)^2+4\left(\frac{6}{5}-6.6461\alpha\right)^2,$$

解得 $\alpha_2=0.09$,置 $x^{(3)}=x^{(2)}+\alpha_2 d^{(2)}=\begin{bmatrix}0\\1.2\end{bmatrix}+0.09\begin{bmatrix}4.4304\\-6.6461\end{bmatrix}=\begin{bmatrix}0.3987\\0.6019\end{bmatrix}$.

第三次迭代:

(2) 确定 $x^{(3)}$ 处的起作用约束指标集

$$I(x^{(3)})=\{1,2\},$$

(3) $l=0$(无等式约束),但 $I(x^{(3)})\neq\varnothing$,则令

$$N^{(3)}=\begin{bmatrix}1&15\\1&10\end{bmatrix},$$

$$\begin{aligned}P^{(3)}&=I-N^{(3)}(N^{(3)T}N^{(3)})^{-1}N^{(3)T}\\&=\begin{bmatrix}1&0\\0&1\end{bmatrix}-\begin{bmatrix}1&15\\1&10\end{bmatrix}\left(\begin{bmatrix}1&1\\15&10\end{bmatrix}\begin{bmatrix}1&15\\1&10\end{bmatrix}\right)^{-1}\begin{bmatrix}1&1\\15&10\end{bmatrix}\end{aligned}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$d^{(3)} = -P^{(3)} \nabla f(x^{(3)})$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

（4）因为 $d^{(3)} = 0$，

(a) $l = 0$，但 $I(x^{(3)}) \neq \varnothing$ 则

$$\lambda^{(3)} = (N^{(3)\mathrm{T}} N^{(3)})^{-1} N^{(3)\mathrm{T}} \nabla f(x^{(3)})$$

$$= \begin{bmatrix} 13 & -1 \\ -1 & 0.08 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 15 & 10 \end{bmatrix} \begin{bmatrix} 0.8 \\ 4.8 \end{bmatrix}$$

$$= \begin{bmatrix} 12.8 \\ -0.8 \end{bmatrix},$$

(b)
$$\lambda_q^{(3)} = \min\{\lambda_i^{(3)} \mid i \in I(x^{(3)})\} = -0.8,$$

置
$$N^{(3)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$P^{(3)} = I - N^{(3)} (N^{(3)\mathrm{T}} N^{(3)})^{-1} N^{(3)\mathrm{T}}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \left([1 \quad 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} [1 \quad 1]$$

$$= \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix},$$

$$d^{(3)} = -P^{(3)} \nabla f(x^{(3)})$$

$$= \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0.8 \\ 4.8 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ -2 \end{bmatrix}.$$

（5）$d^{(3)} \neq 0$，所以，求解一维问题

$$\min \quad \varphi(\alpha) = f(x^{(3)} + \alpha d^{(3)}),$$
$$\text{s.t.} \quad 0 \leqslant \alpha \leqslant \alpha_{\max},$$

其中

$$\alpha_{\max} = \min\left\{ \frac{0 - 0.6}{-2} \right\} = 0.3,$$

即
$$\min_{\alpha \in [0, 0.3]} (0.4 + 2\alpha)^2 + 4(0.6 - 2\alpha)^2,$$

解得 $\alpha_3 = 0.2$，置 $x^{(4)} = x^{(3)} + \alpha_3 d^{(3)} = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} + 0.2 \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix}$.

第四次迭代：

(2) 确定 $x^{(4)}$ 处的起作用约束指标集

$$I(x^{(4)}) = \{1\},$$

(3) $l=0$（无等式约束），但 $I(x^{(4)}) \neq \varnothing$，则令

$$N^{(4)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$P^{(3)} = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix},$$

$$d^{(4)} = -P^{(4)} \nabla f(x^{(4)}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

(4) 因为 $d^{(4)} = \mathbf{0}$，

(a) $l=0$，但 $I(x^{(4)}) \neq \varnothing$ 则

$$\lambda^{(4)} = (N^{(4)\mathrm{T}} N^{(4)})^{-1} N^{(4)\mathrm{T}} \nabla f(x^{(4)}) = 1.6 > 0,$$

(b) $\lambda_q^{(4)} \geqslant 0$，则停止计算，$x^{(4)}$ 即为 K－T 点，所以

$$x^* = x^{(4)} = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix}.$$

14.3 既约梯度法

既约梯度法的基本思想是将求解线性规划的单纯形法推广到求解非线性规划问题（因其是由 Wolfe 提出的，因此也称为 Wolfe 既约梯度法）.

考虑线性约束问题

$$\begin{aligned} &\min f(x), x \in \mathbf{R}^n, \\ &\text{s.t. } Ax = b, \\ &x \geqslant \mathbf{0}, \end{aligned} \tag{14-44}$$

其中 $A = A_{m \times n}(m < n)$，$\mathrm{rank}(A) = m$，假设，$A$ 的任意 m 列均是线性无关的.

若对 A 作分解

$$A = [B, N], \tag{14-45}$$

其中 $B = B_{m \times m}$，是可逆矩阵. 此时相应的分解为

$$x^{\mathrm{T}} = [x_B^{\mathrm{T}}, x_N^{\mathrm{T}}] (x \text{ 是可行点}) \tag{14-46}$$

（其中 $x_B > \mathbf{0}$ 称为基变量，$x_N \geqslant \mathbf{0}$ 称为非基变量），

$$\nabla f(x)^{\mathrm{T}} = (\nabla_B f(x)^{\mathrm{T}}, \nabla_N f(x)^{\mathrm{T}}). \tag{14-47}$$

设 $\boldsymbol{d}$ 是约束问题(14－44)的可行方向，则 $\boldsymbol{d}$ 满足

$$\boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}, \tag{14－48}$$

$$d_j \geqslant 0 (\text{当 } x_j = 0 \text{ 时}), \tag{14－49}$$

则 $\boldsymbol{d}$ 亦可作相应的划分

$$\boldsymbol{d}^{\mathrm{T}} = (\boldsymbol{d}_B^{\mathrm{T}}, \boldsymbol{d}_N^{\mathrm{T}}), \tag{14－50}$$

则有

$$\boldsymbol{B}\boldsymbol{d}_B + \boldsymbol{N}\boldsymbol{d}_N = \boldsymbol{0}, \tag{14－51}$$

上式左乘 $\boldsymbol{B}^{-1}$ 得

$$\boldsymbol{d}_B = -\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{d}_N, \tag{14－52}$$

若 $\boldsymbol{d}$ 再满足

$$\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d} < 0, \tag{14－53}$$

则 $\boldsymbol{d}$ 是 $\boldsymbol{x}$ 处的下降方向，将(14－53)式改写成

$$\nabla_B f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}_B + \nabla_N f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}_N < 0, \tag{14－54}$$

将(14－52)式代入(14－53)式，得

$$[\nabla_N f(\boldsymbol{x})^{\mathrm{T}} - \nabla_B f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{N}]\boldsymbol{d}_N < 0. \tag{14－55}$$

定义 14.5(既约梯度)若令

$$\begin{aligned}
\boldsymbol{r}^{\mathrm{T}} &= (\boldsymbol{r}_B^{\mathrm{T}}, \boldsymbol{r}_N^{\mathrm{T}}) \\
&= \nabla f(\boldsymbol{x})^{\mathrm{T}} - \nabla_B f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{A} \\
&= (\nabla_B f(\boldsymbol{x})^{\mathrm{T}}, \nabla_N f(\boldsymbol{x})^{\mathrm{T}}) - \nabla_B f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{B}, \boldsymbol{N}) \\
&= (\boldsymbol{0}, \nabla_N f(\boldsymbol{x})^{\mathrm{T}} - \nabla_B f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{N}),
\end{aligned} \tag{14－56}$$

则称 $\boldsymbol{r}$ 为目标函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的既约梯度.

由上述推导可知，$\boldsymbol{d}$ 满足

$$\boldsymbol{d}_B = -\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{d}_N, \tag{14－57}$$

$$d_j \geqslant 0 \quad (\text{当 } x_j = 0 \text{ 时}), \tag{14－58}$$

$$\boldsymbol{r}_N^{\mathrm{T}}\boldsymbol{d}_N < 0, \tag{14－59}$$

则 $\boldsymbol{d}$ 是 $\boldsymbol{x}$ 处的可行下降方向，因此，可以按照如下方法选择 $\boldsymbol{d}_N$，令

$$d_j = \begin{cases} -r_j, & r_j \leqslant 0, \\ -x_j r_j, & r_j > 0, \end{cases} \quad j \in \mathbf{N}, \tag{14－60}$$

其中 $\mathbf{N}$ 表示非基变量指标集.

定理 14.8　设 $\boldsymbol{x}$ 是可行点，$f(\boldsymbol{x})$ 具有连续的一阶偏导数，若 $\boldsymbol{x} = [\boldsymbol{x}_B, \boldsymbol{x}_N]^{\mathrm{T}}$，$\boldsymbol{x}_B > \boldsymbol{0}$，$\boldsymbol{A} = [\boldsymbol{B}, \boldsymbol{N}]$ 且 $\boldsymbol{B}^{-1}$ 存在，$\boldsymbol{r}^{\mathrm{T}} = \nabla f(\boldsymbol{x})^{\mathrm{T}} - \nabla_B f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{A}$，$\boldsymbol{d} \in \mathbf{R}^n$ 是按(14－60)式和(14－57)式构造的. 如果 $\boldsymbol{d} \neq \boldsymbol{0}$，则 $\boldsymbol{d}$ 是可行下降方向，进一步 $\boldsymbol{d} = \boldsymbol{0}$ 的充分必要条件是：$\boldsymbol{x}$ 是 K－T 点.

证 显然由(14-60)式构造的 d_j 满足(14-58)式,因此 $\boldsymbol{d}$ 是可行方向.

下证 $\boldsymbol{d}$ 是下降方向(即证明 $\boldsymbol{d}$ 满足(14-59)式). 由(14-60)式得到

$$\begin{aligned} \boldsymbol{r}_N^{\mathrm{T}}\boldsymbol{d}_N &= \sum_{j\in N} r_j d_j = \sum_{r_j\leqslant 0} r_j d_j + \sum_{r_j>0} r_j d_j \\ &= -\sum_{r_j\leqslant 0} r_j^2 - \sum_{r_j>0} x_j r_j^2 \leqslant 0. \end{aligned} \tag{14-61}$$

因为 $\boldsymbol{d}\neq\boldsymbol{0}$, $\boldsymbol{d}_N\neq\boldsymbol{0}$. 因此存在 $r_j<0$ 或 $x_j>0$ 且 $r_j>0$,所以(14-61)式中小于号严格成立.

最后证明定理的第二部分.

问题(14-44)的局部解的 K-T 条件为:存在 $\boldsymbol{u}$、$\boldsymbol{v}$ 使得

$$\nabla f(\boldsymbol{x}) + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{0}, \tag{14-62}$$

$$\boldsymbol{v} \geqslant \boldsymbol{0}, \tag{14-63}$$

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x} = 0 \tag{14-64}$$

对 $\boldsymbol{v}$ 作划分,令 $\boldsymbol{v}^{\mathrm{T}}=(\boldsymbol{v}_B^{\mathrm{T}},\boldsymbol{v}_N^{\mathrm{T}})$,由(14-63)式和 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 可知(14-64)式等价于:

$$\boldsymbol{v}_B^{\mathrm{T}}\boldsymbol{x}_B = 0 \,\&\, \boldsymbol{v}_N^{\mathrm{T}}\boldsymbol{x}_N = 0. \tag{14-65}$$

注意到 $\boldsymbol{x}_B>\boldsymbol{0}$,因此 $\boldsymbol{v}_B=\boldsymbol{0}$. 这样(14-62)式化为

$$\nabla_B f(\boldsymbol{x})^{\mathrm{T}} + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{0}, \tag{14-66}$$

$$\nabla_B f(\boldsymbol{x})^{\mathrm{T}} + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{N} - \boldsymbol{v}_N^{\mathrm{T}} = \boldsymbol{0}. \tag{14-67}$$

由(14-66)式可知,

$$\boldsymbol{u}^{\mathrm{T}} = -\nabla_B f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}^{-1}, \tag{14-68}$$

将(14-68)式代入(14-67)式,得到

$$\boldsymbol{u}_N^{\mathrm{T}} = \nabla_B f(\boldsymbol{x})^{\mathrm{T}} - \nabla_B f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{N} = \boldsymbol{r}_N^{\mathrm{T}}, \tag{14-69}$$

因此,K-T 条件转化为

$$\boldsymbol{r}_N \geqslant \boldsymbol{0}, \tag{14-70}$$

$$\boldsymbol{r}_N^{\mathrm{T}}\boldsymbol{x}_N = 0. \tag{14-71}$$

当 $\boldsymbol{d}=\boldsymbol{0}$ 时,即 $\boldsymbol{d}_N=\boldsymbol{0}$,因此有

$$r_j = 0, \tag{14-72}$$

或

$$x_j r_j = 0, \text{且 } r_j > 0, \tag{14-73}$$

即(14-70)式和(14-71)式成立,因此 $\boldsymbol{x}$ 为 K-T 点.

反过来,若 $\boldsymbol{x}$ 是 K-T 点,则(14-70)式和(14-71)式成立. 由(14-60)式得 $d_j=0$,即 $\boldsymbol{d}_N=\boldsymbol{0}$. 因此 $\boldsymbol{d}_B=\boldsymbol{0}$,从而 $\boldsymbol{d}=\boldsymbol{0}$.

一维搜索

下面讨论既约梯度法中一维搜索的搜索区间,由(14-60)式可知,

当 $x_j=0$ 时，有 $d_j\geqslant 0$，因此不会破坏 x 的可行性. 只有当 $x_j>0$ 且 $d_j<0$ 时，x 的可行性才有可能被破坏. 因此取

$$\alpha_{\max}=\min\left\{-\frac{x_j}{d_j}\,\middle|\,d_j<0,x_j>0\right\},$$

当 $\{d_j<0 \text{ 且 } x_j>0\}=\varnothing$ 时，取 $\alpha_{\max}=+\infty$.

算法 14.5（既约梯度法）

① 取初始可行点 $x^{(1)}$，即 $x^{(1)}$ 满足

$$Ax^{(1)}=b,$$
$$x^{(1)}\geqslant 0,$$

置 $k=1$；

② 构造指标集 $I_k=\{j|x_j \text{ 是 } x^{(k)} \text{ 的 } m \text{ 个最大分量之一}\}$，

并计算

$$B=(a_j|j\in I_k),$$

$$N=(a_j|j\notin I_k),$$

$$r_N^{\mathrm T}=\nabla_N f(x^{(k)})^{\mathrm T}-\nabla_B f(x^{(k)})^{\mathrm T}B^{-1}N,$$

$$d_j^{(k)}=\begin{cases}-r_j, & r_j\leqslant 0,\\ -x_jr_j, & r_j>0,\end{cases}\quad j\notin I_k,$$

$$d_B^{(k)}=-B^{-1}Nd_N^{(k)};$$

③ 若 $d^{(k)}=0$（一般为 $\|d^{(k)}\|\leqslant\varepsilon$），则停止计算，$x^{(k)}$ 为最优解；否则

求解一维问题

$$\min\quad \phi(\alpha)=f(x^{(k)}+\alpha d^{(k)}),$$

$$\text{s.t.}\quad 0\leqslant\alpha\leqslant\alpha_{\max},$$

其中

$$\alpha_{\max}=\begin{cases}\min\left\{\dfrac{-x_j^{(k)}}{d_j^{(k)}}\,\middle|\,d_j^{(k)}<0,x_j^{(k)}>0\right\}\\ +\infty,\ \{d_j^{(k)}<0,x_j^{(k)}>0\}=\varnothing,\end{cases}$$

得到 α_k，置 $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}$；

④ 置 $k=k+1$，转②.

例 14.5 用既约梯度法求解：

$$\begin{cases} \min & (x_1-1)^2+(x_2-2)^2, \\ \text{s.t.} & x_1+x_2+x_3=2, \\ & -2x_1+3x_2+x_4=3, \\ & x_1,x_2,x_3,x_4\geqslant 0. \end{cases}$$

解 (1) 取初始可行点 $x^{(1)}=\begin{bmatrix}0\\0\\2\\3\end{bmatrix}$，$\nabla f(x)=[2(x_1-1),2(x_2-2),0,0]^{\mathrm{T}}$，置 $k=1$；

(2) 确定 $x^{(1)}$ 处的起作用约束指标集

$$I(x^{(1)})=\{3,4\},$$

$$B=(a_3,a_4)=\begin{bmatrix}1&0\\0&1\end{bmatrix},$$

$$N=(a_1,a_2)=\begin{bmatrix}1&1\\-2&3\end{bmatrix},$$

$$r_N^{\mathrm{T}}=[-2\quad -4]-[0\quad 0]\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}1&1\\-2&3\end{bmatrix}$$

$$=[-2\quad -4],$$

$$d_1^{(1)}=2, d_2^{(1)}=4,$$

$$d_B^{(1)}=-\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}1&1\\-2&3\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix}$$

$$=\begin{bmatrix}-6\\-8\end{bmatrix},$$

$$d^{(1)}=[2,4,-6,-8]^{\mathrm{T}}.$$

(3) 显然 $d^{(1)}\neq 0$，则求解一维问题

$$\min \quad \phi(\alpha)=f(x^{(1)}+\alpha d^{(1)}),$$

$$\text{s.t.} \quad 0 \leqslant \alpha \leqslant \alpha_{\max},$$

其中

$$\alpha_{\max} = \min\left\{\frac{-2}{-6}, \frac{-3}{-8}\right\} = \frac{1}{3},$$

即
$$\min_{\alpha \in [0, \frac{1}{3}]} [(2\alpha - 1)^2 + (4\alpha - 2)^2],$$

解得 $\alpha_1 = \frac{1}{3}$，置 $x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = \begin{bmatrix} 2/3 \\ 4/3 \\ 0 \\ 1/3 \end{bmatrix}$.

第二次迭代：

(2) 确定 $x^{(2)}$ 处的起作用约束指标集

$$I(x^{(2)}) = \{1, 2\},$$

$$B = \begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix},$$

$$N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$r_N^{\mathrm{T}} = [0 \quad 0] - \left[-\frac{2}{3} \quad -\frac{4}{3}\right]\begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \left[\frac{14}{15} \quad \frac{2}{15}\right]$$

$$d_3^{(2)} = -x_3 r_3 = 0, \quad d_4^{(2)} = -x_4 r_4 = -\frac{2}{45},$$

$$d_B^{(2)} = -\begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ -\frac{2}{45} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{2}{225} \\ \frac{2}{225} \end{bmatrix},$$

$$d^{(2)} = \left[-\frac{2}{225}, \frac{2}{225}, 0, -\frac{2}{45} \right]^T.$$

（3）因为 $d^{(2)} \neq 0$，则求解一维问题

$$\min \quad \phi(\alpha) = f(x^{(2)} + \alpha d^{(2)}),$$

$$\text{s.t.} \quad 0 \leqslant \alpha \leqslant \alpha_{max},$$

其中

$$\alpha_{max} = \min \left\{ \frac{-\frac{2}{3}}{-\frac{2}{225}}, \frac{-\frac{1}{3}}{-\frac{2}{45}} \right\} = 7.5,$$

即
$$\min_{\alpha \in [0, 7.5]} \left[\left(-\frac{2}{225}\alpha - \frac{1}{3} \right)^2 + \left(\frac{2}{225}\alpha - \frac{2}{3} \right)^2 \right],$$

解得 $\alpha_2 = \frac{15}{2}$，置 $x^{(3)} = x^{(2)} + \alpha_2 d^{(2)} = \begin{bmatrix} 3/5 \\ 7/5 \\ 0 \\ 0 \end{bmatrix}$.

第三次迭代：

（2）确定 $x^{(3)}$ 处的起作用约束指标集

$$I(x^{(2)}) = \{1, 2\},$$

$$B = \begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix},$$

$$N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$r_N^{\mathrm{T}} = [0 \quad 0] - \left[-\frac{4}{5} \quad -\frac{6}{5}\right]\begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \left[\frac{24}{25} \quad \frac{2}{25}\right],$$

$$d_3^{(3)} = -x_3 r_3 = 0, d_4^{(3)} = -x_4 r_4 = 0,$$

$$d_B^{(3)} = -\begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\boldsymbol{d}^{(3)} = \boldsymbol{0}.$$

（3）因为 $\boldsymbol{d}^{(3)} = \boldsymbol{0}$，所以 $\boldsymbol{x}^{(3)}$ 即为 K－T 点，

即
$$\boldsymbol{x}^* = \boldsymbol{x}^{(3)} = \begin{bmatrix} 3/5 \\ 7/5 \\ 0 \\ 0 \end{bmatrix}.$$

第 15 章 罚 函 数 法

本章方法是将约束优化问题变为无约束问题来求解，这便要求无约束问题的目标函数必须是原来的目标函数与约束函数的某种组合（线性规划中，大 M 法求初始基本可行解时，即是如此），罚函数法就是在原来的目标函数上加上由约束函数组成的一个“惩罚”项，来迫使迭代点逼近可行域的.

在以下讨论中，假定所有函数是连续的.

15.1 罚 函 数 法

15.1.1 罚函数法

1. 等式约束下的罚函数法

考虑等式约束问题

$$\begin{aligned} &\min f(\boldsymbol{x}), \\ &\text{s.t. } h_i(\boldsymbol{x}) = 0, i = 1,2,\cdots,l. \end{aligned} \tag{15-1}$$

假定它的最优解存在，可行域为

$$D = \{\boldsymbol{x} \mid h_i(\boldsymbol{x}) = 0, i = 1,2,\cdots,l\},$$

令

$$p(\boldsymbol{x}) = \sum_{i=1}^{l} [h_i(\boldsymbol{x})]^2, \tag{15-2}$$

进而考虑求解新的函数

$$F(\boldsymbol{x},M) = f(\boldsymbol{x}) + Mp(\boldsymbol{x}) = f(\boldsymbol{x}) + M\sum_{i=1}^{l} [h_i(\boldsymbol{x})]^2 \tag{15-3}$$

的无约束极小问题. 其中 M 为一个充分大的正数.

我们希望 $F(\boldsymbol{x},M)$ 的最优解 $\boldsymbol{x}(M)$，当 M 充分大时能逼近约束问题 (15-1) 的最优解. 为此，附加项 $Mp(\boldsymbol{x})$ 应具有这样的性质：当 $\boldsymbol{x} \in D$，即满足约束条件 $h_i(\boldsymbol{x}) = 0, i = 1,2,\cdots,l$ 时，附加项为 0；当 $\boldsymbol{x} \notin D$，即不满足约束条件时，附加项给以很大的惩罚，也即，这个函数 $F(\boldsymbol{x},M)$ 对可行点不实行惩罚，而对非可行点应给予很大的惩罚，即

$$M\sum_{i=1}^{l} [h_i(\boldsymbol{x})]^2 = \begin{cases} 0, & \boldsymbol{x} \in D, \\ > 0, & \boldsymbol{x} \notin D, \end{cases}$$

从而求解等式约束问题(15－1),可转化为求解无约束问题

$$\min_{x \in \mathbf{R}^n} F(x,M).$$

定义 15.1 以上这种将约束极值转化为无约束极值的方法,称为罚函数法. 称 $F(x,M)$ 为罚函数;M 为罚因子;$Mp(x)$ 为罚项.

2. 不等式约束下的罚函数法

同样,对于不等式约束问题

$$\begin{aligned} \min \quad & f(x), x \in \mathbf{R}^n, \\ \text{s.t.} \quad & g_i(x) \geqslant 0, i = 1,2,\cdots,m, \end{aligned} \tag{15-4}$$

由于不等式约束 $g_i(x) \geqslant 0, i = 1,2,\cdots,m$

等价于等式约束

$$\min(0,g_i(x)) = 0, i = 1,2,\cdots,m,$$

因此不等式约束问题(15－4),可转化为等式约束问题

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & \min(0,g_i(x)) = 0, i = 1,2,\cdots,m. \end{aligned} \tag{15-5}$$

记 $\mathrm{D} = \{x \mid \min(0,g_i(x)) = 0, i = 1,2,\cdots,m\}$,

$$p(x) = \sum_{i=1}^{m} [\min(0,g_i(x))]^2,$$

构造函数

$$F(x,M) = f(x) + Mp(x) = f(x) + M\sum_{i=1}^{m} [\min(0,g_i(x))]^2,$$

其中 M 为一个充分大的正数,且

$$M\sum_{i=1}^{m} [\min(0,g_i(x))]^2 = \begin{cases} 0, & x \in \mathrm{D}, \\ > 0, & x \notin \mathrm{D}, \end{cases} \tag{15-6}$$

如此便将求解约束极值(15－5)的问题转化为求解无约束极值问题

$$\min_{x \in \mathbf{R}^n} F(x,M). \tag{15-7}$$

定理 15.1 对某个确定的 $M > 0$,若 $x(M)$ 是无约束极值问题(15－7)的最优解,则 $x(M) \in \mathrm{D}$ 必是约束极值问题(15－5)的最优解.

证 若 $x(M) \in \mathrm{D}$,则 $\sum_{i=1}^{m}[\min(0,g_i(x))]^2 = 0$,所以对任意 $x \in \mathrm{D}$,总有

$$f(x(M)) = F(x(M),M) = \min_{x \in \mathbf{R}^n} F(x,M) \leqslant F(x,M) = f(x),$$

由 x 的任意性知 $x(M) \in \mathrm{D}$ 是问题(15－7)的最优解,定理得证.

3. 罚因子的确定

定理 15.1 说明,只要求出无约束极值问题(15－7)的最优解,且它属于可行域,即 $x(M) \in \mathrm{D}$,便求出了约束极值;若其最优解 $x(M) \notin \mathrm{D}$,

则应增大罚因子 M，再重复上面的讨论．一般取 M 为一个数列 M_k，如取定

$$0 < M_1 < M_2 < \cdots < M_k < \cdots,$$

求解
$$\min_{x \in \mathbf{R}^n} F(x, M_k),$$

若对应于 M_k 的罚函数的最优解 $x^{(k)} = x(M_k) \notin D$，可以证明，$x^{(k)}$ 与可行域 D 的距离将随 k 的增大而越来越小．亦即，$x^{(k)}$ 是从可行域 D 的外部趋向于约束极值(15－5)的最优解的．所以也称为 SUMT 外点法．

4．罚函数算法

算法 15.1（罚函数法）

① 选择序列 $\{M_k\}$，$M_k = cM_{k-1}\quad(c>1)$，$M_k \to +\infty\ (k\to+\infty)$，取初始点 $x^{(0)}$，置精度要求 ε，置 $k=1$；

② 以 $x^{(k-1)}$ 为初始点，求解无约束问题

$$\min_{x \in \mathbf{R}^n} F(x, M_k) = f(x) + M_k \sum_{i=1}^{m} [\min(0, g_i(x))]^2, \quad (15-8)$$

得到最优解 $x^{(k)}$；

③ 若 $M_k p(x^{(k)}) \leqslant \varepsilon$，则停止（$x^{(k)}$ 即可作为约束问题的最优解）；否则置 $k = k+1$，转②．

例 15.1 用罚函数法求解约束问题

$$\min \quad f(x) = \frac{1}{2}x_1^2 + \frac{1}{6}x_2^2,$$
$$\text{s.t.} \quad x_1 + x_2 = 1,$$

其中 $x = (x_1, x_2)^T$，要求选取 $M_1 = 0.05$，$c = 2$，$\varepsilon = 10^{-5}$．

解 引入罚函数

$$F(x, M_k) = \left(\frac{1}{2}x_1^2 + \frac{1}{6}x_2^2\right) + M_k(x_1 + x_2 - 1)^2,$$

求解无约束问题
$$\min_{x \in \mathbf{R}^n} F(x, M_k),$$

令
$$F'_{x_1}(x, M_k) = x_1 + 2M_k(x_1 + x_2 - 1) = 0,$$
$$F'_{x_2}(x, M_k) = \frac{1}{3}x_2 + 2M_k(x_1 + x_2 - 1) = 0,$$

解得
$$x^{(k)} = (x_1^{(k)}, x_2^{(k)})^T = \left(\frac{2M_k}{1+8M_k}, \frac{6M_k}{1+8M_k}\right)^T,$$

$M_k \to +\infty$，得 $x^* = \left(\dfrac{1}{4}, \dfrac{3}{4}\right)^T$．

例 15.2 用罚函数法求解约束问题

$$\min f(x) = x_1^2 + x_2^2,$$

$$\text{s.t.}\quad x_1 \geqslant 1$$

(显然,最优解为 $x^* = (1,0)^T, f(x^*) = 1$).

解　引入罚函数

$$F(x,M) = (x_1^2 + x_2^2) + M[\min(0,x_1 - 1)]^2$$
$$= \begin{cases} x_1^2 + x_2^2, & x_1 \geqslant 1, \\ x_1^2 + x_2^2 + M(x_1 - 1)^2, & x_1 < 1, \end{cases}$$

求解无约束问题

$$\min_{x \in \mathbf{R}^n} F(x, M_k),$$

$$0 = F'_{x_1}(x,M) = \begin{cases} 2x_1, & x_1 \geqslant 1, \\ 2x_1 + 2M(x_1 - 1), & x_1 < 1, \end{cases}$$

令

$$0 = F'_{x_2}(x,M) = 2x_2,$$

解得

$$(x_1(M), x_2(M))^T = \left(\frac{M}{1+M}, 0\right)^T,$$

此时

$$F(x,M) = \left(\frac{M}{1+M}\right)^2 + M\left(\frac{1}{1+M}\right)^2 = \frac{M}{1+M}.$$

令　$M \to +\infty$,得 $x^* = (1,0)^T$,　$f(x^*) = 1$.

15.1.2　罚函数法的收敛性质

定理 15.2　设 $0 < M_k < M_{k+1}$,$x^{(k)}, x^{(k+1)}$ 分别是取罚因子 M_k, M_{k+1} 时的无约束问题(15 - 7)的最优解,则下列各式成立:

(1) $F(x^{(k)}, M_k) \leqslant F(x^{(k+1)}, M_{k+1})$;

(2) $p(x^{(k)}) \geqslant p(x^{(k+1)})$;

(3) $f(x^{(k)}) \geqslant f(x^{(k+1)})$.

证　(1) 因为 $x^{(k)}$ 是问题(15 - 7)取罚因子 M_k 时的最优解,所以

$$F(x^{(k)}, M_k) \leqslant F(x^{(k+1)}, M_k) = f(x^{(k+1)}) + M_k p(x^{(k+1)})$$
$$\leqslant f(x^{(k+1)}) + M_{k+1} p(x^{(k+1)}) = F(x^{(k+1)}, M_{k+1}).$$

(2) 因为 $x^{(k)}, x^{(k+1)}$ 分别是取罚因子 M_k, M_{k+1} 时的无约束问题(15 - 7)的最优解,所以有

$$\begin{aligned} f(x^{(k)}) + M_k p(x^{(k)}) &\leqslant f(x^{(k+1)}) + M_k p(x^{(k+1)}), \\ f(x^{(k+1)}) + M_{k+1} p(x^{(k+1)}) &\leqslant f(x^{(k)}) + M_{k+1} p(x^{(k)}). \end{aligned} \qquad (15 - 9)$$

将上面两式的两端分别相加,经整理后可得

$$(M_{k+1} - M_k)[p(x^{(k)}) - p(x^{(k+1)})] \geqslant 0,$$

又$(M_{k+1}-M_k)>0$，所以得到$p(x^{(k)})\geqslant p(x^{(k+1)})$；

(3) 由(15-9)式，又由 $$p(x^{(k)})\geqslant p(x^{(k+1)}),$$
得 $$f(x^{(k)})-f(x^{(k+1)})\geqslant 0$$
所以 $$f(x^{(k)})\geqslant f(x^{(k+1)}).$$

由上述定理可知，如果迭代不终止，那么$\{f(x^{(k)})\}$和$\{F(x^{(k)},M_k)\}$为单调递增序列，$\{p(x^{(k)})\}$为单调递减序列.

下面给出罚函数法的收敛性定理.

定理 15.3 设(1) 函数$f(x)$，$g_i(x)(i=1,2,\cdots,m)$是$\mathbf{R}^n$上的连续函数；

(2) 序列$\{M_k\}$满足 $$0<M_1<M_2<\cdots<M_k<\cdots,$$
且 $$\lim_{k\to\infty}M_k=+\infty;$$

(3) 对任意$M_k>0$，问题$\min\limits_{x\in\mathbf{R}^n}F(x,M_k)$的最优解$x^{(k)}=x(M_k)$存在且有界；

(4) 约束最优化问题(15-5)存在最优解x^*.

则(a)对某个$k(\geqslant 1)$，若$x^{(k)}\in D$，$x^{(k)}$为问题(15-5)的最优解；

(b)若情形(1)总不发生，则得到点列$\{x^{(k)}\}$，其任意极限点都是问题(15-5)的最优解.

证 (a)显然，由定理15.1结论成立；

下证(b)，由已知，设 $$\lim_{k\to\infty}x^{(k)}=\bar{x},$$
由于$f(x^{(k)})\leqslant f(x^{(k)})+M_kp(x^{(k)})=F(x^{(k)},M_k)=\min\limits_{x\in\mathbf{R}^n}F(x,M_k)$
$\leqslant F(x^*,M_k)=f(x^*)$，
又由f的连续性，令$k\to\infty$得
$$f(\bar{x})\leqslant f(x^*),\tag{15-10}$$
下证$\bar{x}\in D$，由定理15.2得，$\{f(x^{(k)})\}$和$\{F(x^{(k)},M_k)\}$为单调递增序列，由已知问题(15-5)存在最优解x^*，所以$x^*\in D$，从而$p(x^*)=0$，且对任何k有
$$F(x^{(k)},M_k)\leqslant F(x^*,M_k)=f(x^*)+M_kp(x^*)=f(x^*),$$
$$f(x^{(k)})\leqslant f(x^{(k)})+M_kp(x^{(k)})=F(x^{(k)},M_k)\leqslant f(x^*),$$
由此，$\{f(x^{(k)})\}$和$\{F(x^{(k)},M_k)\}$为单调递增且有界的序列，故
$$\lim_{k\to+\infty}F(x^{(k)},M_k)=\bar{F}\leqslant f(x^*),$$
$$\lim_{k\to+\infty}f(x^{(k)})=\bar{f}\leqslant f(x^*),$$
又因为 $$p(x^{(k)})=\frac{F(x^{(k)},M_k)-f(x^{(k)})}{M_k},$$

令 $k\to+\infty$，上式取极限得 $\lim\limits_{k\to+\infty}p(x^{(k)})=0$.

又由 $g_i(x)$ 的连续性得
$$p(\bar{x})=\lim_{k\to+\infty}p(x^{(k)})=0,$$
所以，$\bar{x}\in D$.

又因为 x^* 为最优解，所以
$$f(x^*)\leqslant f(\bar{x}), \tag{15-11}$$
联合(15-10)、(15-11)得出 $f(\bar{x})=f(x^*)$，如此说明 $\bar{x}$ 是问题(15-5)的最优解.

注 1. 罚函数法利用求解无约束极值问题的算法与程序来求解约束极值问题，这是其优点.

2. 但过程中要求一系列无约束极小问题，工作量必然较大，且在求解过程中 M_k 不断增大，当迭代过程收敛时，M_k 常常是很大的，这就可能陷入病态. 且 M_k 很大时，可能导致一个不可行点，或有较大目标函数值的可行点，而这些却是罚函数法的不足之处.

15.2 障碍函数法

15.2.1 算法的构成

障碍函数法的算法思想是，对接近可行域边界的(可行)点施加越来越大的惩罚，对边界上的点施加无穷大的惩罚，以使迭代点封闭在可行域内. 如此便要求可行域的内点集合非空，否则当可行点全被加上无穷大的惩罚时，惩罚则变得毫无意义. 所以，障碍函数法仅适合于不等式约束. 障碍函数法又叫内点法.

考虑不等式约束最优化问题(15-4)
$$\min f(x),\ x\in\mathbf{R}^n,$$
$$\text{s.t. } g_i(x)\geqslant 0, i=1,2,\cdots,m,$$
易知，若可行点 x 是可行域 $D=\{x\mid g_i(x)\geqslant 0, i=1,2,\cdots,m\}$ 的边界点，那么至少存在一个不等式约束，例如第 i 个，使得 $g_i(x)=0$.

不难想到，构造如下的增广目标函数：
$$F(x,\mu)=f(x)+\mu B(x), \tag{15-12}$$
其中
$$B(x)=\sum_{i=1}^{m}\frac{1}{g_i(x)}. \tag{15-13}$$

定义 15.2 如上构造的函数 $B(x)$ 称为障碍函数(内部惩罚函数)，$\mu>0$ 称为惩罚因子；$\mu B(x)$ 称为惩罚项.

若 x 是 D 的内点，惩罚函数 $B(x)$ 的值是有限的正数；当 x 由内部接

近边界时，至少有一个约束函数，例如第 i 个，将使 $g_i(\boldsymbol{x})$ 趋于零，于是 $B(\boldsymbol{x})$ 趋于无穷大，则 $F(\boldsymbol{x},\mu)$ 正是我们希望的函数.

固定 $\mu>0$，求解以(15－12)为目标函数的无约束最优化问题

$$\min F(\boldsymbol{x},\mu), \tag{15-14}$$

如果初始点取为可行域 D 的内点，那么按 $F(\boldsymbol{x},\mu)$ 的结构和无约束下降迭代算法的特性，其迭代点将永远保留在 D 内，最后求得的极小点也必属于 D.

但是，对固定的 $\mu>0$，无约束问题(15－14)的极小点一般不是约束问题(15－4)的极小点. 若取 $\mu>0$ 为趋于零的正数列 $\{\mu_k\}$，则当 $\mu_k\to0$ 时，$\boldsymbol{x}_{\mu_k}\to\boldsymbol{x}^*$. 如此便可通过求解一系列无约束问题来求解约束问题(15－4)，即对 $k=0,1,\cdots$，求解

$$\min\{f(\boldsymbol{x})+\mu_kB(\boldsymbol{x})\},$$

由此得到极小点的序列 $\{\boldsymbol{x}^{(k)}\}$. 后面将证明，若它收敛，则必收敛于约束问题(15－4)的极小点 $\boldsymbol{x}^*$.

注 内部惩罚函数还可有多种取法：

$$B(\boldsymbol{x})=\sum_{i=1}^{m}\frac{1}{[g_i(\boldsymbol{x})]^2},$$

$$B(\boldsymbol{x})=\sum_{i=1}^{m}\ln\frac{1}{g_i(\boldsymbol{x})},$$

$$B(\boldsymbol{x})=\sum_{i=1}^{m}\frac{1}{g_i(\boldsymbol{x})u(-g_i(\boldsymbol{x}))}.$$

其中，$u(t)=\begin{cases}0, & t\geqslant0,\\ 1, & t<0\end{cases}$ 是阶跃函数.

算法 15.2(障碍函数法)

已给定：终止限 ε，惩罚函数

$$B(\boldsymbol{x})=\sum_{i=1}^{m}\frac{1}{g_i(\boldsymbol{x})},$$

① 选定初始可行点 $\boldsymbol{x}^{(0)}$，选取初始惩罚因子 $\mu_1>0$，惩罚因子的缩小系数 $c<1$；置 $k=1$；

② 以 $\boldsymbol{x}^{(k-1)}$ 为初始点，求解无约束问题

$$\min\{f(\boldsymbol{x})+\mu_kB(\boldsymbol{x})\}$$

解得，极小点为 $\boldsymbol{x}^{(k)}$.

③ 若 $\mu_kB(\boldsymbol{x})<\varepsilon$，则 $\boldsymbol{x}^{(k)}$ 就是所求的最优解，停止，否则转④；

④ 置 $\mu_{k+1}=c\mu_k$，$k=k+1$；转②.

例 15.3 用障碍函数法求解

$$\min\left\{\frac{1}{3}x^3\right\},$$

$$\text{s.t.}\ x - a \geqslant 0,$$

其中 a 是非负常数.

解 增广目标函数为

$$F(x,\mu) = \frac{1}{3}x^3 + \frac{\mu}{x - a},$$

用微分法求它的极小点. 令

$$F'_x(x,\mu) = x^2 - \frac{\mu}{(x - a)^2} = 0,$$

由此得极小点所满足的方程为

$$x^2(x - a)^2 = \mu,$$

开方得

$$x(x - a) = \pm\sqrt{\mu},$$

因为要求的 $F(x,\mu)$ 的极小点必须在可行域内,即极小点的坐标满足 $x \geqslant a$,所以舍去上式中的负根 $-\sqrt{\mu}$,得到方程 $x^2 - ax - \sqrt{\mu} = 0$,

其解为
$$x = \frac{1}{2}(a \pm \sqrt{a^2 + 4\sqrt{\mu}}),$$

再舍去负根,最后求得 $F(x,\mu)$ 的极小点为

$$x^{(\mu)} = \frac{1}{2}(a + \sqrt{a^2 + 4\sqrt{\mu}}),$$

令 $\mu \to 0$,求得约束极小点为 $x^* = \lim\limits_{\mu\to 0} x^{(\mu)} = a$,

此题求解过程的几何意义如图 15 - 1 所示.

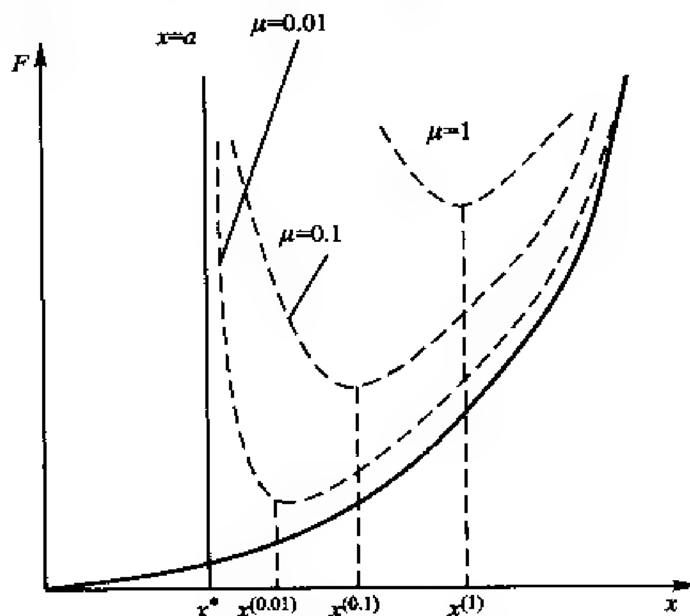


图 15 - 1

15.2.2　障碍函数法的收敛性定理

引理 15.1　对于由算法 15.2 所产生的序列$\{x^{(k)}\}$总有

$$F(x^{(k+1)},\mu_{k+1}) \leqslant F(x^{(k)},\mu_k),$$

其中 $k\geqslant 1$.

证　由(15－12)和$\mu_k>\mu_{k+1}$可得

$$\begin{aligned} F(x^{(k)},\mu_k) &= f(x^{(k)}) + \mu_k B(x^{(k)}) \\ &\geqslant f(x^{(k)}) + \mu_{k+1} B(x^{(k)}) \\ &= F(x^{(k)},\mu_{k+1}), \end{aligned} \tag{15－15}$$

又因为 $x^{(k+1)}$是 $F(x,\mu_{k+1})$的极小点，所以对任意的 x，总有

$$F(x,\mu_{k+1}) \geqslant F(x^{(k+1)},\mu_{k+1}),$$

只需将 x 换为 $x^{(k)}$，即得.

定理 15.4　设$f(x),g_i(x),i=1,2,\cdots,m$ 都是 $\mathbf{R}^n$ 上的连续函数，则由算法 15.2 所产生的任何收敛序列$\{x^{(k)}\}$的极限点必是约束问题(15－4)的极小点.

证　设收敛序列$\{x^{(k)}\}$的极限点是 $\tilde{x}$，约束问题(15－4)的极小点是 x^*. 由增广目标函数 $F(x,\mu)$的定义和$\mu_k B(x^{(k)})\geqslant 0$ 可知，

$$F(x^{(k)},\mu_k) \geqslant f(x^{(k)}) \geqslant f(x^*), \tag{15－16}$$

再由引理得，$\{F(x^{(k)},\mu_k)\}$是有下界的不增序列，因此必有极限.

不妨设

$$\lim_{k\to\infty} F(x^{(k)},\mu_k) = \tilde{F} \geqslant f(x^*), \tag{15－17}$$

若能证明 $\tilde{F}\geqslant f(x^*)$，则由(15－17)和(15－16)可得知

$$\lim_{k\to\infty} f(x^{(k)}) = f(x^*),$$

再由 f 的连续性立刻得到

$$f(\tilde{x}) = \lim_{k\to\infty} f(x^{(k)}) = f(x^*),$$

从而定理得证. 因此我们只须证明，对于充分大的 k，$F(x^{(k)},\mu_k)-f(x^*)$可以任意小.

由 f 的连续性可知，对于任意小的正数 ε，在可行域的内部可以找到一个点 $\bar{x}$，当 $\bar{x}$ 与 x^* 充分接近时总有

$$f(\bar{x}) - f(x^*) < \frac{\varepsilon}{2}, \tag{15－18}$$

再由 $F(x,\mu)$的定义可知，对于可行域的任意内点 x 总有

$$F(x^{(k)},\mu_k) \leqslant f(x) + \mu_k B(x),$$

特别地，$$F(x^{(k)},\mu_k) \leqslant f(\bar{x}) + \mu_k B(\bar{x}),$$
上式的两端同时减去 $f(x^*)$，得
$$F(x^{(k)},\mu_k) - f(x^*) \leqslant f(\bar{x}) - f(x^*) + \mu_k B(\bar{x}). \quad (15-19)$$
因为 $\lim\limits_{k\to\infty}\mu_k = 0$，所以对上述 ε，存在 K，当 $k > K$ 时必有
$$\mu_k B(\bar{x}) < \frac{\varepsilon}{2}, \quad (15-20)$$
联合(15－19)，(15－18)和(15－20)得
$$F(x^{(k)},\mu_k) - f(x^*) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$
定理得证.

我们知道，障碍函数法中必须先知道一个初始内点 $x^{(0)} \in \operatorname{int} D$，在实际问题中，如果初始内点不能凭直观找出时，则必须给出一种求初始内点的算法. 下面介绍一个寻找初始内点的迭代算法.

算法 15.3(求障碍函数法中初始内点的算法)

① 给定初始点 $x^{(0)}$，初始参数 $r_1 > 0$，缩小系数 $\beta \in (0,1)$，令 $k = 0$；

② 确定指标集，令
$$\mathrm{I}_k = \{i \mid g_i(x^{(k)}) = 0, i = 1,2,\cdots,m\},$$
$$\mathrm{J}_k = \{i \mid g_i(x^{(k)}) > 0, i = 1,2,\cdots,m\};$$

③ 检验是否满足终止准则，若 $\mathrm{I}_k = \varnothing$，则 $x^{(k)} \in \operatorname{int} D$，迭代终止；否则，转④；

④ 求解无约束问题. 以 $x^{(k)}$ 为初始点，求无约束问题 $\min\limits_{x\in \mathrm{D}_k} H_{k+1}(x)$ 的最优解 $x^{(k+1)}$，

其中 $$H_{k+1}(x) = -\sum_{i\in \mathrm{I}_k} g_i(x) + r_{k+1}\sum_{i\in \mathrm{J}_k}\frac{1}{g_i(x)},$$
$$\mathrm{D}_k = \{x \mid g_i(x) > 0, i \in \mathrm{J}_k\},$$
令 $r_{k+2} = \beta r_{k+1}$，$k = k+1$，返回②.

15.3　广义 Lagrange 乘子法

算法思想：把罚函数与 Lagrange 函数结合起来，借助于罚函数法的优点，并结合 Lagrange 乘子的性质，构造出更合适的新目标函数，使得在罚因子 M 适当大的情况下就能逐步达到原约束问题的最优解. 由于这种方法既要借助于 Lagrange 乘子迭代进行，而又不同于经典的 Lagrange 乘子法，故称为广义乘子法.

15.3.1 等式约束下的广义乘子法

考虑只带等式约束的最优化问题(15－1)

$$\min \quad f(\boldsymbol{x}),$$
$$\text{s.t.} \quad h_i(\boldsymbol{x}) = 0, i = 1,2,\cdots,l,$$

易知,(15－1)等价于问题

$$\begin{cases}\min \quad [f(\boldsymbol{x}) + \dfrac{M}{2}\sum_{i=1}^{l} h_i^2(\boldsymbol{x})], \\ \text{s.t.} \quad h_i(\boldsymbol{x}) = 0, i = 1,2,\cdots,l,\end{cases} \tag{15－21}$$

其中 $M>0$.

该问题的 Lagrange 函数为

$$L(\boldsymbol{x},M,\boldsymbol{v}) = f(\boldsymbol{x}) + \frac{M}{2}\sum_{i=1}^{l} h_i^2(\boldsymbol{x}) - \sum_{i=1}^{l} v_i h_i(\boldsymbol{x}),$$

由于 $L(\boldsymbol{x},M,\boldsymbol{v})$ 中既有罚项 $\dfrac{M}{2}\sum_{i=1}^{l} h_i^2(\boldsymbol{x})$,又有乘子项 $-\sum_{i=1}^{l} v_i h_i(\boldsymbol{x})$,故称为乘子罚函数.

与罚函数类似,若设$\{M_k\}$为严格单调递增且趋于无穷大的正数列,则我们可以把求解等式约束问题(15－1)转化为求解一系列的无约束问题

$$\min_{x\in\mathbf{R}^n} L(\boldsymbol{x},M_k,\boldsymbol{v}_k) = f(\boldsymbol{x}) + \frac{M_k}{2}\sum_{i=1}^{l} h_i^2(\boldsymbol{x}) - \sum_{i=1}^{l} v_i^{(k)} h_i(\boldsymbol{x}), \tag{15－22}$$

其中 $\boldsymbol{v}_k = (v_1^{(k)},v_2^{(k)},\cdots,v_l^{(k)})^{\mathrm{T}}$ 是第 k 次迭代中采用的 Lagrange 乘子,并且有

定理 15.5 设 $\boldsymbol{x}^{(k)}$ 是无约束问题(15－22)的最优解,则 $\boldsymbol{x}^{(k)}$ 也是问题

$$\min \quad f(\boldsymbol{x}),$$
$$\text{s.t.} \quad h_i(\boldsymbol{x}) = h_i(\boldsymbol{x}^{(k)}), i = 1,2,\cdots,l \tag{15－23}$$

的最优解.

证 由于 $\boldsymbol{x}^{(k)}$ 是无约束问题(15.3.2)的最优解,故对于任意 $\boldsymbol{x}\in\mathbf{R}^n$,都有

$$L(\boldsymbol{x}^{(k)},M_k,\boldsymbol{v}_k) \leqslant L(\boldsymbol{x},M_k,\boldsymbol{v}_k),$$

即

$$f(\boldsymbol{x}^{(k)}) + \frac{M_k}{2}\sum_{i=1}^{l} h_i^2(\boldsymbol{x}^{(k)}) - \sum_{i=1}^{l} v_i^{(k)} h_i(\boldsymbol{x}^{(k)}) \leqslant f(\boldsymbol{x}) +$$

$$\frac{M_k}{2}\sum_{i=1}^{l}h_i^2(\boldsymbol{x}) - \sum_{i=1}^{l}v_i^{(k)}h_i(\boldsymbol{x}),$$

亦即
$$f(\boldsymbol{x}^{(k)}) - f(\boldsymbol{x}) \leqslant \frac{M_k}{2}\sum_{i=1}^{l}[h_i^2(\boldsymbol{x}) - h_i^2(\boldsymbol{x}^{(k)})] - \sum_{i=1}^{l}v_i^{(k)}[h_i(\boldsymbol{x}) - h_i(\boldsymbol{x}^{(k)})].$$

当 $\boldsymbol{x}$ 满足约束条件 $h_i(\boldsymbol{x}) = h_i(\boldsymbol{x}^{(k)}), i = 1,2,\cdots,l$ 时，$f(\boldsymbol{x}^{(k)}) \leqslant f(\boldsymbol{x})$. 所以，$\boldsymbol{x}^{(k)}$ 也是问题(15－23)的最优解.

终止准则的确定：

由定理 15.5 可知，经过适当调整 M_k 与 $\boldsymbol{v}_k$，当

$$h_i(\boldsymbol{x}^{(k)}) = 0, i = 1,2,\cdots,l$$

成立时，$\boldsymbol{x}^{(k)}$ 即为所求等式约束问题(15－1)的最优解. 因此，在下面叙述的算法步骤中，我们将把 $\boldsymbol{x}^{(k)}$ 是否近似地满足该问题的约束条件，即

$$\|\boldsymbol{h}(\boldsymbol{x}^{(k)})\| < \varepsilon$$

作为终止准则，这里 $\varepsilon > 0$ 是预先给定的允许误差，向量函数

$$\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \cdots, h_l(\boldsymbol{x}))^{\mathrm{T}}.$$

M_k 与 $\boldsymbol{v}_k$ 的选取

目前，算法的困难就在于 M_k 与 $\boldsymbol{v}_k$ 的选取. 如果 $f(\boldsymbol{x})$ 和 $h_i(\boldsymbol{x}) = 0$，$i = 1,2,\cdots,l$ 均具有一阶连续偏导数，则

$$\nabla_x L(\boldsymbol{x}, M_k, \boldsymbol{v}_k) = \nabla f(\boldsymbol{x}) - \sum_{i=1}^{l}[v_i^{(k)} - M_k h_i(\boldsymbol{x})] \nabla h_i(\boldsymbol{x}).$$

当给定了充分大的 M_k 并且在 $\boldsymbol{v}_k$ 已知的情况下，求得无约束问题(15－22)的最优解 $\boldsymbol{x}^{(k)}$，则必有

$$\nabla_x L(\boldsymbol{x}^{(k)}, M_k, \boldsymbol{v}_k) = \nabla f(\boldsymbol{x}^{(k)}) - \sum_{i=1}^{l}[v_i^{(k)} - M_k h_i(\boldsymbol{x}^{(k)})] \nabla h_i(\boldsymbol{x}^{(k)}) = 0.$$

因此，为了使 $\boldsymbol{v}_k$ 逼近 Lagrange 乘于 $\boldsymbol{v}$，可按梯度方向建立乘子迭代，即采用

$$v_i^{(k+1)} = v_i^{(k)} - M_k g_i(\boldsymbol{x}^{(k)}), i = 1,2,\cdots,l$$

来修正 $\boldsymbol{v}_k$，在下一次迭代中便求问题

$$\min_{x \in \mathbb{R}^n} L(\boldsymbol{x}, M_k, \boldsymbol{v}_{k+1})$$

的最优解 $\boldsymbol{x}^{(k+1)}$，这样操作下去，直到 $\boldsymbol{v}_k$ 收敛，即直到

$$\|\boldsymbol{h}(\boldsymbol{x}^{(k)})\| < \varepsilon$$

为止. 如果 $|h_i(\boldsymbol{x}^{(k)})|$ 不收敛到 0 或者收敛太慢，就增大 M_k 的值，再进行

迭代(收敛快慢,一般用$\frac{\|\boldsymbol{h}(\boldsymbol{x}^{(k)})\|}{\|\boldsymbol{h}(\boldsymbol{x}^{(k-1)})\|}$来衡量).

综上所述,可给出等式约束下的广义乘子法的算法步骤.

算法 15.4(等式约束下的广义乘子法)

① 给定初始点 $\boldsymbol{x}^{(0)}$,初始乘子 $\boldsymbol{v}_1$,初始罚因子 $M_1>0$,放大系数 $\alpha>1$,允许误差限 $\varepsilon>0$,参数 $\gamma\in(0,1)$,令 $k=1$;

② 求解无约束问题:以 $\boldsymbol{x}^{(k-1)}$ 为初始点,求解无约束问题(15-22),设其最优解为 $\boldsymbol{x}^{(k)}$;

③ 检查终止准则:若 $\|\boldsymbol{h}(\boldsymbol{x}^{(k)})\|<\varepsilon$,则迭代终止,$\boldsymbol{x}^{(k)}$ 为等式约束问题(15-1)的近似最优解;否则,转④;

④ 判断收敛快慢;若
$$\frac{\|\boldsymbol{h}(\boldsymbol{x}^{(k)})\|}{\|\boldsymbol{h}(\boldsymbol{x}^{(k-1)})\|}\geqslant\gamma,$$
则令 $M_{k+1}=\alpha M_k$,转⑤;否则,令 $M_{k+1}=M_k$,转⑤;

⑤ 进行乘子迭代:令
$$v_i^{(k+1)}=v_i^{(k)}-M_k h_i(\boldsymbol{x}^{(k)}),i=1,2,\cdots,l$$
及 $k=k+1$,返回②.

例 15.4　用乘子法求解
$$\begin{cases}\min\ f(x)=x_1^2-3x_2-x_2^2,\\ x_2=0.\end{cases}$$

解　乘子函数为
$$\begin{aligned}L(x,M,v)&=x_1^2-3x_2-x_2^2+\frac{M}{2}(x_2^2)-vx_2\\&=x_1^2-(3+v)x_2+\frac{1}{2}(M-2)x_2^2.\end{aligned}$$

第一次迭代:

(1) 令 $\boldsymbol{x}^{(0)}=(0,0.5)^{\mathrm{T}}$,$v_1=4$,$M_1=8$,$\alpha=10$,$\varepsilon=0.01$,$\gamma=0.25$.

(2) 此时　$L(\boldsymbol{x},8,0)=x_1^2-7x_2+3x_2^2$,
$$\begin{cases}\dfrac{\partial L}{\partial x_1}=2x_1=0,\\ \dfrac{\partial L}{\partial x_2}=-7+6x_2=0,\end{cases}$$
解得
$$\boldsymbol{x}^{(1)}=\left(0,\frac{7}{6}\right)^{\mathrm{T}}.$$

(3) $\|\boldsymbol{h}(\boldsymbol{x}^{(1)})\|=\frac{7}{6}>\varepsilon$,

(4) 因为$\dfrac{\|\boldsymbol{h}(\boldsymbol{x}^{(1)})\|}{\|\boldsymbol{h}(\boldsymbol{x}^{(0)})\|}=\dfrac{\frac{7}{6}}{\frac{1}{2}}=\dfrac{7}{3}\geqslant 0.25$，令 $M_2=10\,M_1=80$，

(5) $v^{(2)}=4-8\times\dfrac{7}{6}=4-\dfrac{28}{3}=-\dfrac{16}{3}$，

第二次迭代：

(2) 此时
$$L\left(\boldsymbol{x},80,-\frac{16}{3}\right)=x_1^2-\frac{7}{3}x_2+39x_2^2,$$
$$\begin{cases}\dfrac{\partial L}{\partial x_1}=2x_1=0,\\[2ex]\dfrac{\partial L}{\partial x_2}=\dfrac{7}{3}+78x_2=0,\end{cases}$$
解得
$$\boldsymbol{x}^{(2)}=(0,-0.03)^{\mathrm{T}}.$$

(3) $\|\boldsymbol{h}(\boldsymbol{x}^{(2)})\|=0.03>\varepsilon$，

(4) 因为$\dfrac{\|\boldsymbol{h}(\boldsymbol{x}^{(2)})\|}{\|\boldsymbol{h}(\boldsymbol{x}^{(1)})\|}=\dfrac{0.03}{7}=0.025<0.25$，令 $M_3=M_2=80$，

(5)
$$v^{(3)}=-\frac{16}{3}-80\times(-0.03)=-5.3+2.4=-2.9.$$

第三次迭代：

(2) 此时
$$L(\boldsymbol{x},80,-2.9)=x_1^2-0.1x_2+39x_2^2,$$
$$\begin{cases}\dfrac{\partial L}{\partial x_1}=2x_1=0,\\[2ex]\dfrac{\partial L}{\partial x_2}=-0.1+78x_2=0,\end{cases}$$
解得
$$\boldsymbol{x}^{(3)}=(0,0.001)^{\mathrm{T}},$$

(3) $\|\boldsymbol{h}(\boldsymbol{x}^{(3)})\|=0.001<\varepsilon$，所以，$\boldsymbol{x}^{(3)}=(0,0.001)^{\mathrm{T}}$ 即为近似最优解（其最优解为 $\boldsymbol{x}^*=(0,0)^{\mathrm{T}}$）.

15.3.2　不等式约束下的广义乘子法

下面考虑带有不等式约束的问题(15－1)和(15－4).

对于只带不等式约束的问题(15－4)，引入附加变量 $\boldsymbol{y}=(y_1,y_2,\cdots,y_m)^{\mathrm{T}}$，则问题(15－4)等价于等式约束问题
$$\begin{cases}\min\quad f(\boldsymbol{x}),\\ \text{s.t.}\quad g_i(\boldsymbol{x})-y_i^2=0,i=1,2,\cdots,m,\end{cases}\tag{15-24}$$
问题(15－24)所对应的广义乘子法中的乘子罚函数为

$$L(\boldsymbol{x},\boldsymbol{y},M_k,\boldsymbol{v}_k) = f(\boldsymbol{x}) + \frac{M_k}{2}\sum_{i=1}^{m}[g_i(\boldsymbol{x}) - y_i^2]^2$$

$$- \sum_{i=1}^{m} v_i^{(k)}[g_i(\boldsymbol{x}) - y_i^2]^2,$$

从而把问题(15－4)转化为求解无约束问题

$$\min_{\boldsymbol{x}\in\mathbf{R}^n} L(\boldsymbol{x},\boldsymbol{y},M_k,\boldsymbol{v}_k), \tag{15－25}$$

为把问题(15－25)化成只关于 $\boldsymbol{x}$ 求极小的问题,我们将先对函数 $L(\boldsymbol{x},\boldsymbol{y},M_k,\boldsymbol{v}_k)$关于 $\boldsymbol{y}$ 求极小,由此解出 $\boldsymbol{y}$,再代入问题(15－25)中.

用配方法将 $L(\boldsymbol{x},\boldsymbol{y},M_k,\boldsymbol{v}_k)$转化为

$$L(\boldsymbol{x},\boldsymbol{y},M_k,\boldsymbol{v}_k) = f(\boldsymbol{x})$$

$$+ \sum_{i=1}^{m}\left\{\frac{M_k}{2}\left[y_i^2 - \frac{1}{M_k}(M_k g_i(\boldsymbol{x}) - v_i^{(k)})\right]^2 - \frac{(v_i^{(k)})^2}{2M_k}\right\}, \tag{15－26}$$

为使 L 关于 y_i 取极小值,则 y_i 应满足

(1) 当 $M_k g_i(\boldsymbol{x}) - v_i^{(k)} \geqslant 0$ 时,$y_i^2 = \frac{1}{M_k}[M_k g_i(\boldsymbol{x}) - v_i^{(k)}]$;

(2) 当 $M_k g_i(\boldsymbol{x}) - v_i^{(k)} < 0$ 时,$y_i^2 = 0$.

综合(1)、(2)即有

$$y_i^2 = \frac{1}{M_k}\max\{0, M_k g_i(\boldsymbol{x}) - v_i^{(k)}\}, i = 1,2,\cdots,m, \tag{15－27}$$

把(15－27)式代入(15－26)式中,由此定义出与 y_i 无关的乘子罚函数

$$L(\boldsymbol{x},M_k,\boldsymbol{v}_k) = f(\boldsymbol{x})$$

$$+ \frac{1}{2M_k}\sum_{i=1}^{m}\{[\max(0,(v_i^{(k)} - M_k g_i(\boldsymbol{x}))]^2 - (v_i^{(k)})^2\},$$

从而,问题(15－4)转化为求解一系列无约束问题

$$\min_{\boldsymbol{x}\in\mathbf{R}^n} L(\boldsymbol{x},M_k,\boldsymbol{v}_k).$$

15.3.3 等式与不等式约束下的广义乘子法

对于既含有不等式约束又含有等式约束的一般最优化极小问题(15－1),可定义乘子罚函数

$$L(\boldsymbol{x},M_k,\boldsymbol{v}_k,\boldsymbol{\mu}_k) = f(\boldsymbol{x}) + \frac{1}{2M_k}\sum_{i=1}^{m}\{[\max(0,(v_i^{(k)} -$$

$$M_k g_i(\boldsymbol{x}))]^2 - (v_i^{(k)})^2\} + \frac{M_k}{2}\sum_{i=1}^{l}h_i^2(\boldsymbol{x}) + \sum_{i=1}^{l}\mu_i^{(k)}h_i(\boldsymbol{x}),$$

其中乘子迭代公式为

$$v_i^{(k+1)} = \max\{0, v_i^{(k)} - M_k g_i(\boldsymbol{x}^{(k)})\}, i = 1,2,\cdots,m,$$

$$\mu_i^{(k+1)} = \max\{0, \mu_i^{(k)} - M_k h_i(x^{(k)})\}, i = 1,2,\cdots,l,$$

因此,问题(15－1)转化为求解一系列无约束问题

$$\min_{x \in \mathbf{R}^n} L(x, M_k, v_k, \mu_k),$$

上述两类问题的具体计算步骤与等式约束的情形类似.

例 15.5　用广义乘子法求解

$$\min \quad (x_1 + 1)^2 + x_2^2,$$

$$\text{s.t.} \quad x_1 + x_2 \geqslant 1,$$

解　增广目标函数为

$$L(x, M, v) = (x_1 + 1)^2 + x_2^2 + \frac{1}{2M}\{[\max(0, v - M(x_1 + x_2 - 1)]^2 - v^2\}$$

$$= \begin{cases} (x_1 + 1)^2 + x_2^2 - \dfrac{v^2}{2M}, & x_1 + x_2 - 1 > \dfrac{v}{M}, \\ (x_1 + 1)^2 + x_2^2 + \dfrac{1}{2M}\{[v - M(x_1 + x_2 - 1)]^2 - v^2\}, \\ x_1 + x_2 - 1 \leqslant \dfrac{v}{M}. \end{cases}$$

当 $x_1 + x_2 - 1 > \dfrac{v}{M}$时,

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 2(x_1 + 1) = 0, \\ \dfrac{\partial L}{\partial x_2} = 2x_2 = 0, \end{cases}$$

解得　　$(x_1, x_2)^{\mathrm{T}} = (-1, 0)^{\mathrm{T}}$,显然不满足约束.

当 $x_1 + x_2 - 1 < \dfrac{v}{M}$时,

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 2(x_1 + 1) + \dfrac{2}{2M}[v - M(x_1 + x_2 - 1)](-M) \\ \qquad = 2x_1 + 2 + M(x_1 + x_2 - 1) - v = 0, \\ \dfrac{\partial L}{\partial x_2} = 2x_2 + \dfrac{2}{2M}[v - M(x_1 + x_2 - 1)](-M) \\ \qquad = 2x_2 + M(x_1 + x_2 - 1) - v = 0, \end{cases}$$

解得　　$(x_1, x_2)^{\mathrm{T}} = \left(\dfrac{v - 2}{2 + 2M}, \dfrac{v + 2M}{2 + 2M}\right)^{\mathrm{T}}$,

由乘子迭代公式

$$v_i^{(k+1)} = \max\{0, v_i^{(k)} - M_k h_i(x^{(k)})\},$$

又结合　　$(x_1, x_2)^{\mathrm{T}} = \left(\dfrac{v - 2}{2 + 2M}, \dfrac{v + 2M}{2 + 2M}\right)^{\mathrm{T}}$,

得 $v^*=2$,所以 $(x_1,x_2)^*=(0,1)^T$,此时 M 可取一固定值.

由于罚因子 M_k 不必趋向无穷大就能求得原约束问题的最优解,因此,避免了外罚函数法中的病态出现,计算经验表明,广义乘子法优于外罚函数法.

15.4 精确罚函数法

精确罚函数法的思想:与罚函数法相比,精确罚函数法不需要求解一系列无约束问题,而是构造一个函数,它不需要依赖某些未知参数,且它的无约束极小点恰好就是原约束问题的最优解.

精确罚函数法包括 Zangwill 的经典精确罚函数法(见文献 5),以及 Fletcher 提出的可微精确罚函数法. 由于篇幅所限此处我们只介绍后者.

15.4.1 非线性等式约束问题的可微精确罚函数微法

问题
$$\begin{aligned}&\min \quad f(\boldsymbol{x}),\\ &\text{s.t.} \quad h_i(\boldsymbol{x})=0, i=1,2,\cdots,l,\end{aligned} \tag{15-28}$$

构造广义 Lagrange 函数
$$L(\boldsymbol{x})=f(\boldsymbol{x})-\boldsymbol{\lambda}(\boldsymbol{x})^T\boldsymbol{h}(\boldsymbol{x}), \tag{15-29}$$
其中 $\boldsymbol{\lambda}(\boldsymbol{x})=[\lambda_1(\boldsymbol{x}),\lambda_2(\boldsymbol{x}),\cdots,\lambda_l(\boldsymbol{x})]^T$,$\boldsymbol{h}(\boldsymbol{x})=[h_1(\boldsymbol{x}),h_2(\boldsymbol{x}),\cdots,h_l(\boldsymbol{x})]^T$.

此广义 Lagrange 函数和经典 Lagrange 函数相比,区别在于:$\boldsymbol{\lambda}$ 不再是一个独立的向量,而是 $\boldsymbol{x}$ 的函数.

由
$$\nabla L(\boldsymbol{x})=\boldsymbol{S}(\boldsymbol{x})-\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x})-[\nabla\boldsymbol{\lambda}(\boldsymbol{x})]\boldsymbol{h}(\boldsymbol{x}), \tag{15-30}$$
其中
$$\boldsymbol{A}(\boldsymbol{x})=\nabla\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{S}(\boldsymbol{x})=\nabla f(\boldsymbol{x}),$$
知,要使 $\nabla L(\boldsymbol{x}^*)=\boldsymbol{0}$,则 $\boldsymbol{h}(\boldsymbol{x}^*)=\boldsymbol{0}$,
且
$$\boldsymbol{S}(\boldsymbol{x}^*)=\boldsymbol{A}(\boldsymbol{x}^*)\boldsymbol{\lambda}(\boldsymbol{x}^*), \tag{15-31}$$
于是构造 $\boldsymbol{\lambda}(\boldsymbol{x})$ 的一个方法是
$$\boldsymbol{\lambda}(\boldsymbol{x})=[\boldsymbol{A}(\boldsymbol{x})]^+\boldsymbol{S}(\boldsymbol{x}), \tag{15-32}$$
这是超定线性方程组(15-31)的极小最小二乘解.

因此
$$L(\boldsymbol{x})=f(\boldsymbol{x})-\boldsymbol{h}(\boldsymbol{x})^T[\boldsymbol{A}(\boldsymbol{x})]^+\boldsymbol{S}(\boldsymbol{x}), \tag{15-33}$$
由矩阵代数知,满足 $\nabla L(\boldsymbol{x}^*)=\boldsymbol{0}$ 的点 $\boldsymbol{x}^*$ 未必是(15-33)的极小点. 然而若在 $L(\boldsymbol{x})$ 中加进具有正定矩阵 $\boldsymbol{Q}$ 的平方项 $\boldsymbol{h}(\boldsymbol{x})^T\boldsymbol{Q}\boldsymbol{h}(\boldsymbol{x})$,则不仅可以保持 $\nabla L(\boldsymbol{x}^*)=\boldsymbol{0}$,且可望在 $\boldsymbol{x}^*$ 处得到正的曲率. 一种特殊的选择是取

$Q=\mu[A(x)^+]\times[A(x)^+]^{\mathrm T}$，于是

$$L(x)=f(x)-h(x)^{\mathrm T}$$
$$[A(x)^+]S(x)+\mu h(x)^{\mathrm T}[A(x)^+][A(x)^+]^{\mathrm T}h(x). \quad (15-34)$$

定理 15.6 设 x^* 是问题(15-28)的严格局部极小点，则对于(15-34)成立 $\nabla L(x^*)=0$，而且存在一个 μ^*，使得对每一个 $\mu\geq\mu^*$，矩阵 $\nabla^2 L(x^*)$ 是正定的(证明略去).

推论 若定理 15.6 的条件成立，则存在一个 μ^*，使得对每一个 $\mu\geq\mu^*$，(15-34)的无约束极小点就是(15-28)的最优点.

由以上分析，可以建立解非线性等式约束问题的精确罚函数算法.

算法 15.5(精确罚函数法) 给定初始点 $x^{(0)}$，置 $r_1=1,k=1$.

① 计算

$$\mu=10\|\nabla^2 f(x_{k-1})\|+r_k; \quad (15-35)$$

② 以 $x^{(k-1)}$ 为初始点，μ 为参数值，求解

$$\min \quad L(x), \quad (15-36)$$

其中 $L(x)$ 由(15-34)确定，得 $x^{(k)}$.

③ 判断 $\nabla^2 L(x^{(k)})$ 的正定性：若 $\nabla^2 L(x^{(k)})$ 正定，则 $x^*=x^{(k)}$，停止；否则，令 $r_{k+1}=1.5r_k$，置 $k=k+1$，转①.

15.4.2 一般非线性约束问题的可微精确罚函数法

若令 $w(x)=(g(x),h(x))^{\mathrm T}$，则(15-34)还可以表示成

$$L(x)=f(x)-[r(x)]^{\mathrm T}w(x), \quad (15-37)$$

其中 $r(x)=[A(x)]^+S(x)-\mu[A(x)]^+[[A(x)]^+]^{\mathrm T}w(x)$，

对于 $\hat{x}\in\mathbf{R}^n$，引入记号：$A(\hat{x})=\hat{A},g(\hat{x})=\hat{g},w(\hat{x})=\hat{w}$，考虑二次规划问题

$$\begin{aligned}\min \quad & \frac{1}{2}\mu\delta^{\mathrm T}\delta+g\delta,\\ \text{s.t.} \quad & \hat{A}\delta+\hat{w}=0,\end{aligned} \quad (15-38)$$

便可导出如下的结论.

定理 15.7 在问题(15-38)中，若 $\mathrm{rank}(\hat{A})=l$，λ^* 为其 Lagrange 乘子，则 $\lambda^*=r(\hat{x})$.

证 构造问题(15-38)的 Lagrange 函数

$$\phi(\delta,\lambda)=\frac{1}{2}\mu\delta^{\mathrm T}\delta+\hat{g}^{\mathrm T}\delta-\lambda^{\mathrm T}(\hat{A}^{\mathrm T}\delta+\hat{w}),$$

由 $\nabla_\delta\phi(\delta,\lambda)=0$，得

$$\lambda^* = \hat{A}^+ (\mu\delta + \hat{s}),$$

又因 $\delta = -(\hat{A}^T)+\hat{w}$,所以

$$\lambda^* = \hat{A} + \hat{s} - \mu\hat{A}^+ (\hat{A}^+)^T\hat{w},$$

即 $\lambda^* = r(\hat{x})$.

定理 15.7 构成了解(NP)问题可微精确罚函数法的基础. 在(NP)问题中,对于 $\hat{x} \in \mathbf{R}^n$,引入记号:

$$\hat{s} = \nabla f(\hat{x}), \hat{A}_1 = \nabla g(\hat{x})(g_i(x) \geqslant 0),$$

$$\hat{A}_2 = \nabla h(\hat{x})(h_i(x) = 0), \hat{w} = w(\hat{x}).$$

考虑二次规划问题

$$\begin{aligned} \min \quad & \frac{1}{2}\mu\delta^T\delta + \hat{s}^T\delta, \\ \text{s.t.} \quad & \hat{A}_1^T\delta + \hat{w} \geqslant 0, \\ & \hat{A}_2^T\delta + \hat{w} \geqslant 0. \end{aligned} \tag{15-39}$$

假如问题(15-39)解点处的有效约束也是(NP)问题解点处的有效约束,并且(NP)问题和(15-39)的 Lagrange 乘子相同,那么,按照问题(15-39)的 Lagrange 乘子向量 u^*, v^*,就可以构造惩罚函数

$$L(x) = f(x) - (u^*)^T w(x), \tag{15-40}$$

解无约束问题

$$\min \quad L(x), \tag{15-41}$$

就可以得到(NP)的最优解 x^*.

精确罚函数法计算 $L(x)$时,不但要求 $f(x), w(x)$,且要求 $\nabla f(x)$, $\nabla w(x)$,特别在无约束极值求解时,一旦要求 $\nabla L(x)$,则进一步要求 $\nabla^2 f(x)$,所以非常烦琐,正因为如此,精确罚函数法很少应用于实际中.

第 16 章 二 次 规 划

16.1 二次规划问题及其 K－T 条件

二次规划问题形式可表示为：

（一）
$$\min \quad Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x},$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$
$$\boldsymbol{x} \geqslant \boldsymbol{0},$$

其中 $\boldsymbol{H}$ 是 n 阶对称半正定矩阵，$\boldsymbol{p}$ 是 n 维列向量，$\boldsymbol{A}$ 是 $m \times n$ 矩阵，$\boldsymbol{b}$ 是 m 维列向量.

（二）
$$\min \quad Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x},$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b},$$
$$\boldsymbol{x} \geqslant \boldsymbol{0},$$

其中 $\boldsymbol{H}$ 是 n 阶对称半正定矩阵，$\boldsymbol{p}$ 是 n 维列向量，$\boldsymbol{A}$ 是 $m \times n$ 矩阵，$\boldsymbol{b}$ 是 m 维列向量.

（三）
$$\min \quad Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x},$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b},$$
$$\boldsymbol{x} \geqslant \boldsymbol{0},$$

其中 $\boldsymbol{H}$ 是 n 阶对称半正定矩阵，$\boldsymbol{p}$ 是 n 维列向量，$\boldsymbol{A}$ 是 $m \times n$ 矩阵，$\boldsymbol{b}$ 是 m 维列向量.

二次规划在实际应用中很重要，另外二次规划解法也经常是解一般非线性约束最优化问题的工具，故此二次规划的计算方法越来越引起人们的注意. 下面我们简要介绍二次规划的 K－T 条件以便为后面方法的讲述作基础.

不失一般性，设二次规划问题为

$$\min \quad Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x},$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leqslant \boldsymbol{0},$$
$$\boldsymbol{x} \geqslant \boldsymbol{0}.$$

引入 Lagrange 函数

$$L(\boldsymbol{x},\boldsymbol{u}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{u}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}),$$

由第十二章的结论,在适当条件下规划问题的 K－T 条件可写成:

$$\nabla_x L(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{p} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} \geqslant \boldsymbol{0},$$

$$\boldsymbol{x}^{\mathrm{T}} \nabla_x L(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}) = 0,$$

$$\boldsymbol{x} \geqslant \boldsymbol{0},$$

$$\nabla_u L(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leqslant \boldsymbol{0},$$

$$\boldsymbol{u}^{\mathrm{T}} \nabla_u L(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{u}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = 0,$$

$$u_i \geqslant 0.$$

现引入 $\boldsymbol{v}\geqslant\boldsymbol{0},\boldsymbol{y}\geqslant\boldsymbol{0}$,上面各式可写成

$$\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{0},$$

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v} = 0,$$

$$\boldsymbol{x}' \geqslant \boldsymbol{0},$$

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} + \boldsymbol{y} = \boldsymbol{0},$$

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{y} = 0,$$

$$\boldsymbol{u} \geqslant \boldsymbol{0},$$

又由 $\boldsymbol{x}\geqslant\boldsymbol{0},\boldsymbol{v}\geqslant\boldsymbol{0},\boldsymbol{y}\geqslant\boldsymbol{0},\boldsymbol{u}\geqslant\boldsymbol{0}$,上式可写成

(1) $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{b}$,

(2) $\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{0}$,

(3) $\boldsymbol{x} \geqslant \boldsymbol{0},\boldsymbol{v} \geqslant \boldsymbol{0},\boldsymbol{y} \geqslant \boldsymbol{0},\boldsymbol{u} \geqslant \boldsymbol{0}$,

(4) $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v} + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{y} = 0$.

为了使用方便,后面再给出另外二种形式二次规划问题的 K－T 条件.

对于二次规划问题(一),K－T 条件为

(1) $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$,

(2) $\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{0}$,

(3) $\boldsymbol{x} \geqslant \boldsymbol{0},\boldsymbol{v} \geqslant \boldsymbol{0}$,

(4) $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v} = 0$.

对于二次规划问题(三),K－T 条件为

(1) $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{b}$,

(2) $\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{0}$,

(3) $\boldsymbol{u} \geqslant \boldsymbol{0},\boldsymbol{y} \geqslant \boldsymbol{0}$,

(4) $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{y} = 0$.

明显的,这些条件都是相应二次规划解的必要条件.

有关充分条件,我们仅就问题(二)进行讨论.

定理 16.1 若对满足 K-T 条件的点(x^*,v^*,y^*,u^*),满足

$$Q(x)-Q(x^*)\geqslant(x-x^*)^T\nabla Q(x^*),\qquad(16-1)$$

则 x^* 便是二次规划的最优解.

证

$$\begin{aligned}Q(x)-Q(x^*)&\geqslant(x-x^*)^T\nabla Q(x^*)\\&=(x-x^*)^T(Hx^*+p)\\&=(x-x^*)^T(v^*-A^Tu^*)\\&=x^Tv^*-x^{*T}v^*+u^{*T}[Ax^*-b-(Ax-b)]\\&=x^Tv^*-x^{*T}v^*-u^{*T}y^*+u^{*T}y\\&\geqslant x^Tv^*-u^{*T}y^*\\&=0,\end{aligned}$$

即对任何满足

$$Ax\leqslant b,$$
$$x\geqslant 0$$

的 x 均有 $$Q(x)\geqslant Q(x^*),$$

所以 x^* 是二次规划的最优解.

注 对二次规划问题(二),因为 $H(x)$是半正定矩阵,则 $Q(x)=\frac{1}{2}x^THx+p^Tx$是凸函数,又由 $Q(x)$的凸性可推出(16-1)成立.

H 为半正定矩阵,则 K-T 条件便是二次规划最优解的充分条件. 这是许多二次规划计算方法的理论基础.

16.2 Lemke 算法

Lemke 算法的基本思想是,把线性规划的单纯形方法加以适当修改,用来求二次规划的 K-T 点.

考虑二次规划问题

$$\begin{aligned}\min\quad &Q(x)\triangleq\frac{1}{2}x^THx+p^Tx,\\ \text{s.t}\quad &Ax\geqslant b,\qquad(16-2)\\ &x\geqslant 0,\end{aligned}$$

其中 H 是 n 阶对称矩阵,p 是 n 维列向量,A 是 $m\times n$ 矩阵,秩为 m,b 是 m 维列向量.

引入 Lagrange 函数

$$L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{u}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}, \quad (16-3)$$

再引入松弛变量 $\boldsymbol{y} \geqslant \boldsymbol{0}$，使

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{y}. \quad (16-4)$$

根据第一节结论，可将问题(16－2)的 K－T 条件写成

$$\begin{aligned} \boldsymbol{v} - \boldsymbol{H}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} &= \boldsymbol{p}, \\ \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} &= -\boldsymbol{b}, \\ \boldsymbol{v}^{\mathrm{T}}\boldsymbol{x} &= 0, \\ \boldsymbol{y}^{\mathrm{T}}\boldsymbol{u} &= 0, \\ \boldsymbol{u},\boldsymbol{v},\boldsymbol{x},\boldsymbol{y} &\geqslant \boldsymbol{0}. \end{aligned} \quad (16-5)$$

若记

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{y} \end{bmatrix}, \quad \boldsymbol{z} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{bmatrix},$$

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{H} & -\boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A} & \boldsymbol{O} \end{bmatrix}, \quad \boldsymbol{q} = \begin{bmatrix} \boldsymbol{p} \\ -\boldsymbol{b} \end{bmatrix},$$

于是，(16－5)可写成下列形式：

$$\left.\begin{aligned} \boldsymbol{w} - \boldsymbol{M}\boldsymbol{z} &= \boldsymbol{q} \\ \boldsymbol{w},\boldsymbol{z} &\geqslant \boldsymbol{0} \end{aligned}\right\} \quad (16-6)$$

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{z} = 0, \quad (16-7)$$

其中 $\boldsymbol{w},\boldsymbol{q},\boldsymbol{z}$ 均为 $m+n$ 维列向量，$\boldsymbol{M}$ 则是 $m+n$ 阶矩阵. (16－6)和(16－7)称为线性互补问题，它的每一个解$(\boldsymbol{w},\boldsymbol{z})$具有这样的特征：解的 $2(m+n)$ 个分量中，至少有 $m+n$ 个取零值，而且其中每对变量(w_i,z_i)中至少有一个为零，其余分量均是非负数. 下面研究怎样求出线性互补问题的解.

定义 16.1 若$(\boldsymbol{w},\boldsymbol{z})$是(16－6)的一个基本可行解，且每个互补变量对$(w_i,z_i)$中有一个变量是基变量，则称$(\boldsymbol{w},\boldsymbol{z})$是互补基本可行解.

这样，求二次规划 K－T 点的问题就转化为求互补基本可行解. 现在介绍求互补基本可行解的 Lemke 方法. 分两种情形讨论：

情形 1. 如果 $\boldsymbol{q} \geqslant \boldsymbol{0}$，则$(\boldsymbol{w},\boldsymbol{z}) = (\boldsymbol{q},\boldsymbol{0})$就是一个互补基本可行解.

情形 2. 如果 $\boldsymbol{q} < \boldsymbol{0}$，则引入人工变量 z_0，令

$$\boldsymbol{w} - \boldsymbol{M}\boldsymbol{z} - \boldsymbol{e}z_0 = \boldsymbol{q}, \quad (16-8)$$

$$\boldsymbol{w},\boldsymbol{z},z_0 \geqslant \boldsymbol{0}, \quad (16-9)$$

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{z} = 0, \quad (16-10)$$

其中 $\boldsymbol{e} = (1,\cdots,1)^{\mathrm{T}}$ 是分量全为 1 的 $m+n$ 维列向量.

为求解情形 2，引入准互补基本可行解的概念.

定义 16.2　设(w,z,z_0)是情形 2 的一个可行解，并满足下列条件：

(1) (w,z,z_0)是(16－8)和(16－9)的一个基本可行解.

(2) 对某个 $s\in\{1,\cdots,m+n\}$，w_s 和 z_s 都不是基变量.

(3) z_0 是基变量，每个互补变量对(w_i,z_i)　$(i=1,\cdots,m+n,i\neq s)$中，恰有一个变量是基变量，

则称(w,z,z_0)为准互补基本可行解.

下面用主元消去法求准互补基本可行解.

首先，令

$$z_0=\max\{-q_i\mid i=1,\cdots,m+n\}=-q_k,$$

$$z=\mathbf{0},\quad w=q+ez_0=q-eq_k,$$

则(w,z,z_0)是一个准互补基本可行解，其中 $w_i(i\neq k)$和 z_0 是基变量，其余变量为非基变量. 以(w,z,z_0)为初始解，用主元消去法求新的准互补基本可行解，力图用这种方法迫使 z_0 变为非基变量. 为保持可行性，选择主元时要遵守两条规则：

1. 若 w_l(或 z_l)出基，则 z_l(或 w_l)入基；

2. 按照单纯形方法中的 θ 规则确定出基变量，这样就能实现从一个准互补基本可行解到另一个准互补基本可行解的转换，直到得到互补基本可行解，即 z_0 变为非基变量，或者由情形 2 所定义的可行域无界的结论.

算法 16.1(Lemke 算法)

① 若 $q\geqslant\mathbf{0}$，则停止计算，$(w,z)=(q,\mathbf{0})$是互补基本可行解；否则，用表格形式表示方程组(16－8)，设

$$-q_k=\max\{-q_i\mid i=1,\cdots,m+n\},$$

取 k 行为主行，z_0 对应的列为主列，进行主元消去，令 $y_k=z_k$.

② 设在现行表中变量 y_k 下面的列为 d_k. 若 $d_k\leqslant\mathbf{0}$，则停止计算，得到(16－8)和(16－9)的可行域的极方向；否则，按 θ 规则确定指标 r，使

$$\frac{\bar{q}_r}{d_{rk}}=\min\left\{\frac{\bar{q}_i}{d_{ik}}\,\middle|\,d_{ik}>0\right\},$$

如果 r 行的基变量是 z_0，则转步 4；否则，进行步 3；

③ 设 r 行的基变量为 w_l 或 z_l(对于某个 $l\neq k$)，变量 y_k 入基，以 r 行为主行，y_k 对应的列为主列，进行主元消去. 如果出基变量是 w_l，则令$y_k=z_l$；如果出基变量是 z_l，则 $y_l=w_l$，转步 2；

④ 变量 y_k 入基，z_0 出基. 以 r 行为主，y_k 对应的列为主列，进行主元消去，得到互补基本可行解，停止计算.

例 16.1　用 Lemke 算法求解：

$$\min \quad x_1^2 - x_1x_2 + 2x_2^2 - x_1 - 10x_2,$$
$$\text{s.t.} \quad -3x_1 - 2x_2 \geqslant -6,$$
$$x_1, x_2 \geqslant 0.$$

解 其中

$$H = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}, \quad p = \begin{bmatrix} -1 \\ -10 \end{bmatrix},$$

$$A = [-3, -2], \quad b = -6,$$

$$M = \begin{bmatrix} H & -A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 3 \\ -1 & 4 & 2 \\ -3 & -2 & 0 \end{bmatrix},$$

$$q = \begin{bmatrix} 0 \\ -b \end{bmatrix} - \begin{bmatrix} -1 \\ -10 \\ 6 \end{bmatrix}.$$

线性互补问题是：

$$w_1 - 2z_1 + z_2 - 3z_3 = 1,$$
$$w_2 + z_1 - 4z_2 - 2z_3 = -10,$$
$$w_3 + 3z_1 + 2z_2 = 6,$$
$$w_i \geqslant 0, \ z_i \geqslant 0, \ i = 1,2,3,$$
$$w_iz_i = 0, \quad i = 1,2,3.$$

引入人工变量 z_0，建立下表（表中，[ ]内的数字为主元素）：

	w_1	w_2	w_3	z_1	z_2	z_3	z_0	q
w_1	1	0	0	-2	1	-3	-1	-1
w_2	0	1	0	1	-4	-2	[-1]	-10
w_3	0	0	1	3	2	0	-1	6

$q_k = -10$，z_0 入基，w_2 出基，经主元消去，迭代得

	w_1	w_2	w_3	z_1	z_2	z_3	z_0	$\bar{q}$
w_1	1	-1	0	-3	[5]	-1	0	9
z_0	0	-1	0	-1	4	2	1	10
w_3	0	-1	1	2	6	2	0	16

$y_k = z_2, r = 1$，迭代得

	w_1	w_2	w_3	z_1	z_2	z_3	z_0	$\bar{q}$
z_2	$\frac{1}{5}$	$-\frac{1}{5}$	0	$-\frac{3}{5}$	1	$-\frac{1}{5}$	0	$\frac{9}{5}$
z_0	$-\frac{4}{5}$	$-\frac{1}{5}$	0	$\frac{7}{5}$	0	$\frac{14}{5}$	1	$\frac{14}{5}$
w_3	$-\frac{6}{5}$	$\frac{1}{5}$	1	$\left[\frac{28}{5}\right]$	0	$\frac{16}{5}$	0	$\frac{26}{5}$

$y_k = z_1, r = 3$，主元 $d_{34} = \frac{28}{5}$，迭代得

	w_1	w_2	w_3	z_1	z_2	z_3	z_0	$\bar{q}$
z_2	$\frac{1}{14}$	$-\frac{5}{28}$	$\frac{3}{28}$	0	1	$\frac{1}{7}$	0	$\frac{33}{14}$
z_0	$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{4}$	0	0	[2]	1	$\frac{3}{2}$
z_1	$-\frac{3}{14}$	$\frac{1}{28}$	$\frac{5}{28}$	1	0	$\frac{4}{7}$	0	$\frac{13}{14}$

$y_k = z_2, r = 1$，主元 $d_{26} = 2$，迭代得

	w_1	w_2	w_3	z_1	z_2	z_3	z_0	$\bar{q}$
z_2	$\frac{7}{28}$	$-\frac{9}{56}$	$\frac{7}{56}$	0	1	0	$-\frac{1}{14}$	$\frac{9}{4}$
z_3	$-\frac{1}{4}$	$-\frac{1}{8}$	$-\frac{1}{8}$	0	0	1	$\frac{1}{2}$	$\frac{3}{4}$
z_1	$-\frac{1}{14}$	$\frac{3}{28}$	$\frac{1}{4}$	1	0	0	$-\frac{2}{7}$	$\frac{1}{2}$

由于 $z_0 = 0$，得到互补基本可行解

$$(w_1, w_2, w_3, z_1, z_2, z_3) = \left(0,0,0,\frac{1}{2},\frac{9}{4},\frac{3}{4}\right)^{\mathrm{T}},$$

因此得到 K－T 点 $(x_1, x_2) = \left(\frac{1}{2},\frac{9}{4}\right)^{\mathrm{T}}$.

由于是凸规划，因此 K－T 点也是最优解.

16.3　Wolfe 方法

考虑的问题

$$\min \quad Q(x) = \frac{1}{2}x^{T}Hx + p^{T}x,$$
$$\text{s.t.} \quad Ax = b,$$
$$x \geqslant 0,$$

其中,矩阵 H 是正半定的,$b \geqslant 0$.

易知,问题等价于求出一对向量 x, v 及另一向量 u,使它们满足

$$\begin{aligned} Ax &= b, \\ Hx + A^{T}u - v &= -p, \\ v \geqslant 0, \quad x &\geqslant 0, \\ v^{T}x &= 0, \end{aligned} \tag{16-11}$$

其中最后一个条件称为互补松弛条件. 去掉这个互补松弛限制及变量非负约束之后,所余的问题是一个 $m+n$ 阶 $m+2n$ 维的线性方程组. 下面我们在此形式下求出新的一组线性互补问题的解.

Wolfe 方法分为二种形式:短形式及长形式.

1. Wolfe 方法短形式:此形式适用于 H 为正定矩阵(或者 $p=0$).

引入人工变量,(16-11)可改写成

$$Ax + w = b,$$
$$Hx + A^{T}u - v + z_1 - z_2 = -p,$$
$$v \geqslant 0,\ x \geqslant 0, z_1 \geqslant 0,\ z_2 \geqslant 0,\ w \geqslant 0,$$

其中 x, v 须满足 $$v^{T}x = 0.$$

容易看出只要取人工变量为

$$w_i = b_i,\ i = 1,2,\cdots,m,$$
$$z_1 = (z_1^{(1)}, z_2^{(1)}, \cdots, z_n^{(1)})^{T},$$
$$z_2 = (z_1^{(2)}, z_2^{(2)}, \cdots, z_n^{(2)})^{T},$$
$$\begin{cases} z_i^{(1)} = -p_i, z_i^{(2)} = 0, \text{若 } -p_i \geqslant 0, \\ z_i^{(1)} = 0, z_i^{(2)} = p_i, \text{若 } -p_i < 0, \end{cases} i = 1,2,\cdots,m,$$

其他变量都为 0,便可得到初始基本可行解.

因为 w 与 z_1, z_2 在约束条件中是交错开的,所以我们可以分两个阶段进行.

第一阶段:令 $u=0, v=0$(则此时 $v^{T}x=0$)得如下线性规划问题:

$$\min \quad s_1 = \sum_{j=1}^{m} w_j,$$

$$\text{s.t.}\begin{cases}\boldsymbol{Ax}+\boldsymbol{w}=\boldsymbol{b},\\ \boldsymbol{Hx}+\boldsymbol{z}_1-\boldsymbol{z}_2=-\boldsymbol{p},\\ \boldsymbol{x}\geqslant\boldsymbol{0},\boldsymbol{z}_1\geqslant\boldsymbol{0},\boldsymbol{z}_2\geqslant\boldsymbol{0},\boldsymbol{w}\geqslant\boldsymbol{0},\end{cases}$$

易知,只要问题有可行解,则必有 $s_1^*=0$,这样第一阶段便得出了

$$\begin{gathered}\boldsymbol{Ax}=\boldsymbol{b},\\ \boldsymbol{Hx}+\boldsymbol{z}_1-\boldsymbol{z}_2=-\boldsymbol{p},\\ \boldsymbol{x}\geqslant\boldsymbol{0},\ \boldsymbol{z}_1\geqslant\boldsymbol{0},\ \boldsymbol{z}_2\geqslant\boldsymbol{0}\end{gathered}$$

的一个基可行解.

第二阶段：令 $\boldsymbol{w}=\boldsymbol{0}$,得规划问题：

$$\begin{gathered}\min\quad s_2=\sum_{j=1}^{n}\left(z_j^{(1)}+z_j^{(2)}\right),\\ \boldsymbol{Ax}=\boldsymbol{b},\\ \boldsymbol{Hx}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}-\boldsymbol{v}+\boldsymbol{z}_1-\boldsymbol{z}_2=-\boldsymbol{p},\\ \boldsymbol{v}\geqslant\boldsymbol{0},\ \boldsymbol{x}\geqslant\boldsymbol{0},\ \boldsymbol{z}_1\geqslant\boldsymbol{0},\ \boldsymbol{z}_2\geqslant\boldsymbol{0},\ \boldsymbol{u}\geqslant\boldsymbol{0},\\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}=0.\end{gathered}$$

显然,第一阶段为第二阶段的问题提供了初始基可行解.第二阶段在迭代中只须注意 x_i,v_i 两者只能有一个是基变量就保证了 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}=0$ 的互补松弛条件.

2. Wolfe 方法长形式

此形式解决一般的二次规划问题.

Wolfe 方法的长形式求解分三个阶段:首先令 $\boldsymbol{p}=\boldsymbol{0}$,使问题适合于应用 Wolfe 方法短形式的形状,用短形式的第一、二阶段得出下述问题的一组解.

当 $\boldsymbol{p}=\boldsymbol{0}$ 时,(16－11)式化为

$$\begin{gathered}\boldsymbol{Ax}=\boldsymbol{b},\\ \boldsymbol{Hx}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}-\boldsymbol{v}=\boldsymbol{0},\\ \boldsymbol{v}\geqslant\boldsymbol{0},\ \boldsymbol{x}\geqslant\boldsymbol{0},\\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}=0.\end{gathered}$$

第三阶段：问题是

$$\begin{gathered}\min\quad s_3=-\mu,\\ \boldsymbol{Ax}=\boldsymbol{b},\\ \boldsymbol{Hx}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}-\boldsymbol{v}+\mu\boldsymbol{p}=\boldsymbol{0},\\ \boldsymbol{v}\geqslant\boldsymbol{0},\ \boldsymbol{x}\geqslant\boldsymbol{0},\ \mu\geqslant 0\\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}=0.\end{gathered}\tag{16-12}$$

在第三阶段迭代中我们仍使用单纯形方法，而且保持 x_i, v_i 中只能有一个是基变量，于是得到了一系列点. 可以证明：只要(16-11)有解，则对任何迭代次数 k，一定有 $\sigma_k = \min_j \{\sigma_j\} < 0$. 其中，$\sigma_j$ 是单纯形法的判别数.

因此只有两种可能：(1) 出现半射线情况而停止；(2) 只要对问题加以摄动，不出现(1)的情况，就一定有 $\mu \geq 1$.

定理 16.2　当出现以上(1)、(2)两种情形时我们都可求出当 $\mu = 1$ 时(16-12)的解，从而求出二次规划问题的 K-T 点.

证　对于迭代点 x_r, v_r, u_r, μ_r 易有 $\sigma_k < 0$，于是以 k 为主元列号进行迭代一定有

$$\mu_{r+1} > \mu_r,$$

于是我们迭代中产生了点列

$$x_r, v_r, u_r, \mu_r, \quad r = 1, 2, \cdots,$$

$$0 = \mu_0 < \mu_1 < \mu_2 < \cdots,$$

若在迭代 s 次处得到了半射线情形，则对一切 $\lambda \geq 0$，一定有

$$x_s + \lambda x_s,\ v_s + \lambda v_s,\ u_s + \lambda u_s,\ \mu_s + \lambda \mu_s$$

都满足(16-12). 我们取

$$\lambda = \frac{1 - \mu_s}{\tilde{\mu}_s},$$

便得所证. 注意，其中 $\tilde{x}_s, \tilde{v}_s, \tilde{u}_s, \tilde{\mu}_s$ 构成了极方向.

如不产生半射线情况，总可以有 $\mu_s \geq 1$. 于是取 $\mu_k < 1$，令

$$\alpha = \frac{\mu_s - 1}{\mu_s - \mu_k}, \quad \beta = \frac{1 - \mu_k}{\mu_s - \mu_k},$$

不难证明有

$$\alpha \geq 0, \beta \geq 0, \alpha + \beta = 1,$$

$$\alpha\mu_k + \beta\mu_s = 1,$$

而且　　$\alpha x_k + \beta x_s,\ \alpha v_k + \beta v_s,\ \alpha u_k + \beta u_s,\ \alpha\mu_k + \beta\mu_s$，

一定满足(16-12)，定理得证.

16.4　序列二次规划法

序列二次规划算法是将二次规划问题的求解方法推广应用于求解一般非线性规划问题的一种序列寻优方法.

序列二次规划算法的基本思想是：在每一迭代点 $x^{(k)}$，构造一个二次

规划子问题，以这个子问题的解，作为迭代的搜索方向 $\boldsymbol{d}_k$，并沿该方向作一维搜索，即

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}_k,$$

得 $\boldsymbol{x}^{(k+1)}$. 重复上述迭代过程，直至点列 $\{\boldsymbol{x}^{(k+1)} (k=0,1,2,\cdots)\}$ 最终逼近原问题的近似约束最优点 $\boldsymbol{x}^*$.

对于具有线性约束的非线性规划问题：

$$\begin{aligned} &\min f(x), \\ &\boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b}, \end{aligned} \tag{16 - 13}$$

若再一迭代点 $\boldsymbol{x}^{(k)}$ 处，对目标函数 $f(x)$ 作 Taylor 展开，对其截断到二阶，得

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^{(k)}) + \nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^{(k)}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{H}_k(\boldsymbol{x} - \boldsymbol{x}^{(k)}),$$

其中，

$$\boldsymbol{H}_k = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_1^2} & \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_1 \partial \boldsymbol{x}_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_1 \partial \boldsymbol{x}_n} \\ \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_2 \partial \boldsymbol{x}_1} & \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_2 \partial \boldsymbol{x}_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_n \partial \boldsymbol{x}_1} & \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_n \partial \boldsymbol{x}_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x}^{(k)})}{\partial \boldsymbol{x}_n^2} \end{bmatrix}$$

为点 $\boldsymbol{x}^{(k)}$ 处的 Hesse 矩阵.

这样我们便得到了问题(16 - 13)的在迭代点 $\boldsymbol{x}^{(k)}$ 处的一个近似二次规划问题

$$\begin{aligned} \min \quad & Q(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y} + \boldsymbol{y}^{\mathrm{T}} \nabla f(\boldsymbol{x}^{(k)}), \\ \text{s.t.} \quad & \boldsymbol{a}_i^{\mathrm{T}} y \geqslant 0, \quad i \in \mathrm{I}_k, \\ & |y_i| \leqslant \delta, \quad i = 1,2,\cdots,n, \end{aligned}$$

其中　$\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}^{(k)}$, $\mathrm{I}_k = \{i \mid \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}^{(k)} = b_i, \quad i = 1,2,\cdots,m\}$,

利用前面的方法求解上述的二次规划问题，得其最优解 $\boldsymbol{y}^*$

算法 16.2

① 置 $k=0$;

② 求 $\mathrm{I}_k = \{i \mid \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}^{(k)} = b_i, i=1,2,\cdots,m\}$;

③ 若 $\mathrm{I}_k = \varnothing$，则令 $-\nabla f(\boldsymbol{x}^{(k)}) = \boldsymbol{d}_k$，转⑤；否则，转④；

④ 求解子二次规划：

$$\min \quad Q(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y} + \boldsymbol{y}^{\mathrm{T}} \nabla f(\boldsymbol{x}^{(k)}),$$

$$\text{s.t.}\quad a_i^T y \geqslant 0,\quad i \in I_k,$$
$$|y_i| \leqslant \delta,\quad i = 1,2,\cdots,n,$$

得其最优解 y^*;

⑤ 若 $\|d_k\| \leqslant \varepsilon$,则 $x^* = x^{(k)}$,停止;否则,令 $x^{(k+1)} = x^{(k)} + u_k d_k$,$u_k = \min\{u_1, u_2\}$,且 $f(x^{(k)} + u_1 d_k) = \min_u f(x^{(k)} + u d_k)$,$u_2 = \min\{\beta_i\}$,s. t. $a_i^T(x^{(k)} + \beta_i y) = b_i$,$i \in I_k$,其中 $I_k = \{i \mid a_i^T y < 0, i \in \{1,2,\cdots,m\}/I_k\}$,令 $k = k+1$,转②.

对于非线性规划问题:

$$\begin{cases} \min f(x),\ x \in \mathbf{R}^n, \\ \text{s.t.}\ h_i(x) = 0,\ i = 1,2,\cdots,l, \\ g_i(x) \geqslant 0,\ i = 1,2,\cdots,m. \end{cases}$$

由其算法思想可知,序列二次规划算法的关键是构造并求解原非线性约束问题的一系列二次规划子问题. 而在迭代点 $x^{(k)}$ 构造的子问题可为:

$$\begin{cases} \min[\nabla f(x^{(k)})]^T d + \frac{1}{2} d^T H_k d, \\ \text{s.t.}\ h_i(x^{(k)}) + [\nabla h_i(x^{(k)})]^T d = 0,\ i = 1,2,\cdots,l, \\ g_i(x^{(k)}) + [\nabla g_i(x^{(k)})]^T d \geqslant 0,\ i = 1,2,\cdots,m, \end{cases}$$

式中,搜索方向 $d = [d_1, d_2, \cdots, d_n]^T$ 为变量,$\nabla f(x^{(k)})$,$\nabla h_i(x^{(k)})$,$g_i(x^{(k)})$,H_k 均为确定的量. 因此,该子问题是一个以 d 为变量的二次规划问题,其中的 H_k 是 Lagrange 函数

$$L(x,u,v) = f(x) - \sum_{i=1}^{l} v_i h_i(x) - \sum_{i=1}^{m} u_i g_i(x),$$

在点 $x^{(k)}$ 的 Hesse 矩阵,即有

$$H_k = \nabla_x^2 L(x^{(k)}, u, v)$$
$$= \nabla^2 f(x^{(k)}) - \sum_{i=1}^{l} v_i \nabla^2 h_i(x^{(k)}) - \sum_{i=1}^{m} u_i \nabla^2 g_i(x^{(k)}),$$

由于按此式计算 Hesse 矩阵 H_k 非常困难,故序列二次规划算法是采用前面的变尺度法逐次由变尺度矩阵 A_k 构造变尺度矩阵 A_{k+1} 来逼近 Hesse 矩阵 H_k 的. 其迭代公式为

$$A_{k+1} = A_k + \Delta A_k,$$

其中,ΔA_k 为校正矩阵,且有

$$\Delta A_k = \frac{A_k \Delta x^{(k)} [\Delta x^{(k)}]^T A_k}{[\Delta x^{(k)}]^T A_k \Delta x^{(k)}} + \frac{\Delta S^{(k)} [\Delta S^{(k)}]^T}{[\Delta S^{(k)}]^T \Delta S^{(k)}}.$$

式中，
$$\Delta \boldsymbol{x}^{(k)}=\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)},$$
$$\Delta \boldsymbol{S}^{(k)}=\nabla_x L(\boldsymbol{x}^{(k+1)},\boldsymbol{\lambda}_{k+1})-\nabla_x L(\boldsymbol{x}^{(k)},\boldsymbol{\lambda}_k),$$
迭代开始时，一般可取 $\boldsymbol{A}_0=\boldsymbol{I}$.

因序列二次规划法中采用了变尺度方法构造 Hesse 矩阵，故也称之为约束变尺度法，同时也具备了变尺度法的优点.

第 17 章　离散系统的动态规划方法

20 世纪 50 年代，为了解决多阶段决策过程的最优化问题，美国数学家贝尔曼（Bellman）等人提出了解决这类问题的动态规划方法. 多阶段决策问题是一类离散系统的控制问题. 半个世纪以来，动态规划在理论、方法和应用上均取得了新的进展，成为解决动态优化问题的重要方法.

多阶段决策过程是指某一过程按时间或空间顺序分为若干阶段，而决策序列是在这些阶段中依次做出的. 例如经济发展规划的制定分成各个时期，而每个时期的目标、措施等是依次做出的，使最终目标达到最优.

图 17 - 1 就是一个多阶段（n 段）决策过程，易知，当各个时期的控制规律 $u_i(i=1,2,\cdots)$ 给出后，状态规律 $x_i(i=1,2,\cdots)$ 就随之确定了.

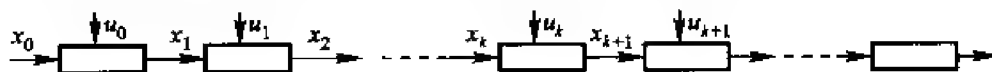


图 17 - 1

离散系统过程显然是一个多阶段决策过程，而用动态规划求解这种决策过程的基本思想是最优性原理. 本章通过两个引例（最短路问题和多阶段资源分配问题）分析多阶段决策系统的基本概念和求解方法.

17.1　多阶段决策问题（引例及相关基本概念）

例 17.1　最短路问题

从 A_0 出发到达终点 A_4，中间必须经过三个阶段，如图 17 - 2 所示.

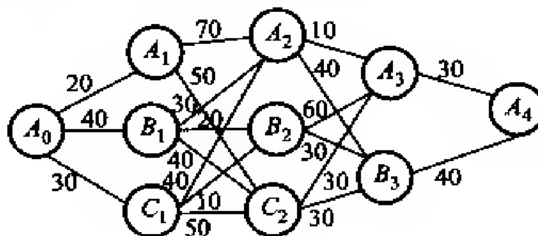


图 17 - 2

第一段在 A_1, B_1, C_1 中任选一个；第二段在 A_2, B_2, C_2 中任选一个；第三段在 A_3, B_3 中任选一个. 图中数字表示连接两点之长，两点间无连线表

示它们之间不能相通. 问题是:选择一条从 A_0 到 A_4 的路径,使其总长为最短.

这是一个定期多阶段决策问题,即阶段数固定,如本例中为四阶段组成的网络.

初态: 以变量 x_0 表示,含一个状态 A_0;

第一段: 以变量 x_1 表示,含 A_1,B_1,C_1 三个状态;

第二段: 以变量 x_2 表示,含 A_2,B_2,C_2 三个状态;

第三段: 以变量 x_3 表示,含 A_3,B_3 两个状态;

第四段(终态): 以变量 x_4 表示,含一个状态 A_4.

下面介绍几个相关的基本概念:

1. 状态变量: 记为 x_k,它表示第 k 段的终点,又是第 $k+1$ 段的起点. 其取值集合包含有限个状态. 如 $\{A_k,B_k,C_k,\cdots,F_k\}$ 在例 17.1 中:

x_0(网络的初态,第一段的起点)取值 A_0;

x_1(第一段的终点,第二段的起点)取值集合为 $\{A_1,B_1,C_1\}$;

x_2(第二段的终点,第三段的起点)取值集合为 $\{A_2,B_2,C_2\}$;

x_3(第三段的终点,第四段的起点)取值集合为 $\{A_3,B_3\}$;

x_4(第四段的终点,亦即网络的终态)取值 A_4.

2. 决策变量:对状态 x_k 的决策称为决策变量,记为 $u_k(x_k)$. 即是从 x_k 转移至 x_{k+1} 所作的选择.

例如:$u_0(x_0)=u_0(A_0)$ 取值 A_1,B_1,C_1

而因为　$u_1(A_1)$ 取值 A_2,C_2;$u_1(B_1)=u_1(C_1)$ 取值 A_2,B_2,C_2.

故所以　$u_1(x_1)$ 既可取值 A_2,C_2,也可以取值 A_2,B_2,C_2.

可见对于同一阶段的状态 A_1,B_1,C_1 可做出不同的决策.

3. 容许决策集合:即决策变量取值的全体. 对于状态变量 x_k,其容许决策集合记为 $D_k(x_k)$.

例如,因为 $D_1(A_1)=\{A_2,C_2\}$, $D_1(B_1)=\{A_2,B_2,C_2\}$, $D_1(C_1)=\{A_2,B_2,C_2\}$,所以 $D_1(x_1)=\{A_2,C_2;A_2,B_2,C_2\}$.

4. 策略:对于一个含有 n 个阶段的决策问题,任何一个由初始状态转移到终止状态的选择,称为全过程策略,简称策略. 记为 $p_{0,n}(x_0)$ 或 $p_{0,n}$. 它是由 n 个决策组成的决策序列,即

$$p_{0,n}=\{u_0(x_0),u_1(x_1),\cdots,u_{n-1}(x_{n-1})\},$$

在图 17-2 中,$p_{0,4}(x_0)=\{A_1,A_2,A_3,A_4\}$,或者 $p_{0,4}(x_0)=\{B_1,A_2,B_3,A_4\},\cdots$. 可见从初态 x_0 到终态 x_4 的转移有多种选择,即是说有多种策略.

5. 子策略:对于一个含有 n 个阶段的决策问题,从初态 x_0 到任何一

个中间状态 x_k 的转移，或从任何一个中间状态到终态的转移所做出的选择为子过程策略，简称子策略. 分别记为

$$p_{0,k}(x_0) = \{u_0(x_0), u_1(x_1), \cdots, u_{k-1}(x_{k-1})\},$$
$$p_{k,n}(x_k) = \{u_k(x_k), u_{k+1}(x_{k+1}), \cdots, u_{n-1}(x_{n-1})\}.$$

6. 容许策略集合：可选择的策略的全体. 记为

$P_{0,k}(x_0) = \{D_0(x_0), D_1(x_1), \cdots, D_{n-1}(x_{n-1})\}$. 例如

$$P_{0,2}(x_0) = \{A_1, A_2; A_1, C_2; B_1, A_2; B_1, B_2; B_1, C_2; C_1, A_2; C_1, B_2; C_1, C_2\}.$$

7. 指标函数：衡量策略优劣的数量指标，记为

$J(x_0, p_{0,n})$——从 x_0 出发，采取策略 $p_{0,n}$ 到终态 x_n 的指标值；

$J(x_0, p_{0,k})$——从 x_0 出发，采取子策略 $p_{0,k}$ 到第 k 段终端某状态的指标值；

$J(x_k, p_{k,n})$——从 x_k 出发，采取子策略 $p_{k,n}$ 到终态 x_n 的指标值；

$L(x_k, u_k)$——从 x_k 出发，采取决策 u_k，使系统的状态转移到 x_{k+1} 的某一个状态的指标值.

在图 17－2 中，$J(A_0, p_{0,4}) = J(A_0, A_1, A_2, A_3, A_4) = 130$；

$J(A_0, p_{0,4}) = J(A_0, B_1, A_2, B_3, A_4) = 150$；

$J(A_0, p_{0,4}) = J(A_0, B_1, A_2, A_3, A_4) = 110$.

可见从初态 A_0 出发，采取策略 $p_{0,4}$ 向终态 A_4 转移，由于 $p_{0,4}$ 可取不同的决策序列，从而得到不同的指标值.

8. 最优指标及最优策略：指标函数的最优值称为最优指标；使指标函数取得最优值的策略称为最优策略. 即对 $\forall p_{k,n}$，若存在 $p_{k,n}^*$ 使得

$$J(x_k, p_{k,n}^*) \leqslant J(x_k, p_{k,n}),$$

称 p^* 为最优策略；$J(x_k, p^*)$ 为最优指标.

在图 17－2 中，由于

$$J(A_0, B_1, A_2, A_3, A_4) = J(A_0, C_1, A_2, A_3, A_4) = J(A_0, C_1, B_2, B_3, A_4) = 110,$$

因此，$\{A_0, B_1, A_2, A_3, A_4\}$，$\{A_0, C_1, A_2, A_3, A_4\}$，$\{A_0, C_1, B_2, B_3, A_4\}$ 均为最优策略，它们共有相同的最优指标值 110.

9. 状态方程：由第 k 阶段的状态变量和决策变量 u_k 完全确定了第 $k+1$ 阶段的状态 x_{k+1}，系统状态的这种转移规律讲为

$$x_{k+1} = T_k(x_k, u_k), \qquad (17-1)$$

称式(17－1)为状态转移方程，简称状态方程. 在例 17.1 中，状态转移规律即状态方程由网络图确定，不能用形如(17－1)的方程表示. 下面再举一个可用解析式表达状态方程的例子.

例 17.2　多阶段资源分配问题

设某工厂有 x_0 台机器生产 A、B 两种产品. 投入 y 台机器生产 A 种产品,纯收入为 $g(y)$(万元);投入 y 台机器生产 B 种产品,纯收入为 $h(y)$(万元),且 $g(0)=h(0)=0$. 又设生产 A、B 两种产品,机器的年折损率分别为 $1-a(a\in(0,1))$ 和 $1-b(b\in(0,1))$. 问在 5 年内如何安排各年度的生产计划,才能使总收入最高?

上述问题可分为 5 个阶段($k=0,1,2,3,4$). 第 k 阶段是从第 k 年末到第 $k+1$ 年安排生产 A 种产品的机器数量.

容许决策集　$\boldsymbol{D}_k(x_k)=\{u_k \mid 0\leqslant u_k\leqslant x_k\}$.

状态方程为　$x_{k+1}=au_k+b(x_k-u_k)=bx_k+(a-b)u_k$.

$$(17-2)$$

17.2　多阶段决策问题的数学描述

17.2.1　数学模型

设有一个由 n 个阶段组成的动态系统,其状态方程为

$$x_{k+1}=T_k(x_k,u_k),\qquad k=0,1,\cdots,n-1,$$

其中 x_0 为初态,x_n 为终态.

给定指标函数 $J(x_0,\boldsymbol{p}_{0,n})$,从容许策略集合 $\boldsymbol{P}_{0,n}(x_0)$ 中确定最优策略

$$\boldsymbol{p}_{0,n}^*(x_0)=\{u_0^*(x_0),u_1^*(x_1),\cdots,u_{n-1}^*(x_{n-1})\}.$$

1) 实现状态转移:即使 x_k 从 x_0 达到 x_n;

2) 使指标函数 $J(x_0,\boldsymbol{p}_{0,n})$ 达到最优值,即

$$J(x_0,\boldsymbol{p}_{0,n}^*)=\underset{p_{0,n}\in P_{0,n}(x_0)}{\mathrm{opt}}J(x_0,\boldsymbol{p}_{0,n}),$$

其中 $J(x_0,\boldsymbol{p}_{0,n})$ 可以表示为阶段效应之和,即

$$J(x_0,\boldsymbol{p}_{0,n})=\sum_{j=0}^{n-1}L(x_j,u_j),$$

其中 $L(x_j,u_j)$ 是系统转移到某一个 x_{j+1} 的指标值. 我们称指标函数 $J(x_0,\boldsymbol{p}_{0,n})$ 具有可加性. 本章仅讨论指标函数具有可加性的多阶段决策问题.

17.2.2　Bellman 最优性原理

在例 17.1 中,利用反证法易证明,若 $A_0\to B_1\to A_2\to B_3\to A_4$ 是一条由 A_0 到 A_4 的最短路,则 $B_1\to A_2\to B_3\to A_4$,$A_2\to B_3\to A_4$ 分别是由 B_1 到 A_4,由 A_2 到 A_4 的最短路. 上述事实正是 20 世纪 50 年代美国数学家 Bellman 等

人提出的最优性原理.

定理 17.1(最优性必要条件)　在一个具有 n 个阶段的决策问题中,若存在容许策略 $p_{0,n}^* = \{u_0^*, u_1^*, \cdots, u_{n-1}^*\}$ 满足

$$J(x_0, p_{0,n}^*) = \underset{p_{0,n} \in P_{0,n}(x_0)}{\mathrm{opt}} J(x_0, p_{0,n}), \qquad (17-3)$$

则对于 $\forall k(1 \leqslant k \leqslant n-1)$,$p_{0,n}^*$ 的子策略 $p_{k,n}^* = \{u_k^*, \cdots, u_{n-1}^*\}$ 必满足

$$J(x_k^*, p_{k,n}^*) = \underset{p_{k,n} \in P_{k,n}(x_k^*)}{\mathrm{opt}} J(x_k^*, p_{k,n}), \qquad (17-4)$$

其中 x_k^* 是由 $p_{0,n}^*$ 确定的第 k 段的状态.

证　不妨假设 opt 为 min. 设相反,$J(x_k^*, p_{k,n}^*) \neq \min\limits_{p_{k,n} \in P_{k,n}} J(x_k^*, p_{k,n})$,

则必存在　$\bar{p}_{k,n} = \{\bar{u}_k \quad \bar{u}_{k+1}, \cdots, \bar{u}_{n-1}\} \in P_{k,n}(x_k^*)$,

使得　$J(x_k^*, \bar{p}_{k,n}) < J(x_k^*, p_{k,n}^*)$.

令　$\hat{p} = \{u_0^*, u_1^*, \cdots, u_{k-1}^*, \bar{u}_k, \bar{u}_{k+1}, \cdots, \bar{u}_{n-1}\}$,

则

$$\begin{aligned} J(x_0, \hat{p}) &= J(x_0, p_{0,k}^*) + J(x_k^*, \bar{p}_{k,n}) \\ &< J(x_0, p_{0,k}^*) + J(x_k^*, p_{k,n}^*) = J(x_0, p_{0,n}^*), \end{aligned}$$

这与 $p_{0,n}^*$ 为全过程最优策略矛盾,即(17-3)式不成立. 从而定理结论成立.

17.2.3　动态规划基本定理

定理 17.2(最优性充分条件)　在一个具有 n 阶段的决策问题中,容许策略 $p_{0,n}^* = \{u_0^*, u_1^*, \cdots, u_{n-1}^*\}$ 是最优策略的充分条件是,对于 $\forall k(1 \leqslant k \leqslant n-1)$ 有

$$J(x_0, p_{0,n}^*) = \underset{p_{0,n} \in P_{0,n}(x_0)}{\mathrm{opt}} \{J(x_0, p_{0,k}) + \mathrm{opt}\, J(\hat{x}_k, p_{k,n})\}, \qquad (17-5)$$

其中 $\hat{x}_k$ 是由 $p_{0,k}$ 确定的第 k 段某一个终端状态.

证　不妨假设 **opt** 为 **min**.

必要性:因为 $p_{0,n}^*$ 是全过程最优策略,所以对于 $\forall p_{0,n} \in P_{0,n}$ 都有

$$J(p_{0,n}^*) \leqslant J(p_{0,n}),$$

对于 $\forall k(1 \leqslant k \leqslant n-1)$,设 $P'_{0,n}$ 是能使由第 k 阶段的终端开始到终止状态的指标取最小值的容许策略集合. 对 $\forall p'_{0,n} \in P'_{0,n}$,都有

$$\begin{aligned} J(p_{0,n}^*) &\leqslant J(x_0, p'_{0,n}) = J(x_0, p'_{0,k}) + J(\hat{x}_k, p'_{k,n}) \\ &= J(x_0, p'_{0,k}) + \min J(\hat{x}_k, p_{k,n}), \end{aligned} \qquad (17-6)$$

式中 $p'_{0,k}$ 和 $p'_{k,n}$ 是 $p'_{0,n}$ 的子策略. 由(17-6)又可得

$$J(p_{0,n}^*) \leqslant \min_{p_{0,k} \in P'_{0,k}(x_0)} \{J(x_0, p'_{0,k}) + \min_{p_{k,n} \in P_{k,n}(\hat{x}_k)} J(\hat{x}_k, p_{k,n})\}, \qquad (17-7)$$

因为 $P'_{0,k}(x_0)=P_{0,k}(x_0)$，故(17－7)式又可写成

$$J(p^*_{0,n}) \leqslant \min_{p_{0,k}\in P_{0,k}(x_0)}\{J(x_0,p_{0,k})+\min_{p_{k,n}\in P_{k,n}(\hat{x}_k)} J(\hat{x}_k,p_{k,n})\}. \tag{17-8}$$

另一方面，对于最优策略 $p^*_{0,n}$，由定理 17.1，应有

$$\begin{aligned}J(p^*_{0,n}) &= J(x_0,p^*_{0,k})+J(\hat{x}_k,p^*_{k,n})\\ &= J(x_0,p^*_{0,k})+\min_{p_{k,n}\in P_{k,n}(\hat{x}_k)} J(\hat{x}_k,p_{k,n}).\end{aligned} \tag{17-9}$$

由于 $p^*_{0,k}\in P_{0,k}(x_0)$，由(17－9)立即得

$$J(p^*_{0,n}) = \min_{p_{0,k}\in P_{0,k}(x_0)}\{J(x_0,p_{0,k})+\min_{p_{k,n}\in P_{k,n}(\hat{x}_k)} J(\hat{x}_k,p_{k,n})\},$$

即(17－5)式成立.

充分性：设 $\bar{p}=\{\bar{u}_0,\bar{u}_1,\cdots,\bar{u}_{n-1}\}$ 是任一容许策略. 对 $\forall k(1\leqslant k\leqslant n-1)$，记

$$\bar{p}_{0,k}=\{\bar{u}_0,\bar{u}_1,\cdots,\bar{u}_{k-1}\},\bar{p}_{k,n}=\{\bar{u}_k,\cdots,\bar{u}_{n-1}\}.$$

设 $\bar{x}_k$ 是由 $\bar{p}_{0,k}$ 所确定的第 k 阶段终态，则

$$J(\bar{x}_k,\bar{p}_{k,n}) \geqslant \min_{p_{k,n}\in P_{k,n}(\bar{x}_k)} J(\bar{x}_k,p_{k,n}),$$

显然 $\bar{p}_{0,k}(x_0)\subset P_{0,k}(x_0)$，从面

$$\begin{aligned}J(\bar{p}) &= J(x_0,\bar{p}_{0,k})+J(\hat{x}_k,p_{k,n})\\ &\geqslant J(x_0,\bar{p}_{0,k})+\min_{p_{k,n}\in P_{k,n}(\bar{x}_k)} J(\bar{x}_k,p_{k,n})\\ &\geqslant \min_{p_{0,k}\in P_{0,k}(x_0)}\{J(x_0,p_{0,k})+\min_{p_{k,n}\in P_{k,n}(\bar{x}_k)} J(\bar{x}_k,p_{k,n})\}\\ &\geqslant \min_{p_{0,k}\in P_{0,k}(x_0)}\{J(x_0,p_{0,k})+\min_{p_{k,n}\in P_{k,n}(\hat{x}_k)} J(\hat{x}_k,p_{k,n})\},\end{aligned}$$

由(17－5)式，有 $J(\bar{p})\geqslant J(p^*)$. 据 $\bar{p}$ 的任意性，由此可知 $p^*_{0,n}$ 必为最优策略.

如果我们记

$$V_k(x_k) = \mathop{\mathrm{opt}}_{p_{k,n}\in P_{k,n}(x_k)} J(x_k,p_{k,n}),\qquad k=0,1,\cdots,n-1,$$

$$V_n(x_n)=0,$$

则由定理 17.2 立即得到如下推论

推论 17.1(动态规划基本方程)　对于一个 n 段决策问题，对 $\forall k(1\leqslant k\leqslant n-1)$，

$$V_k(x_k) = \mathop{\mathrm{opt}}_{u_k\in D_k(x_k)}\{L(x_k,u_k)+V_{k+1}(\hat{x}_{k+1})\}, \tag{17-10}$$

其中 $\hat{x}_{k+1}=T_k(x_k,u_k)$.

$V_k(x_k)$ 称为最优值函数，它表示从 x_k 的任一状态，取最优策略 $p^*_{k,n}(x_k)$

转移到终点 x_n 的最优指标.

$V_{k+1}(\hat{x}_{k+1})$ 表示从 x_{k+1} 的某一状态 $\hat{x}_{k+1}$，取最优策略 $p^*_{k+1,n}(x_{k+1})$ 转移到终点 x_n 的指标.

$L(x_k,u_k)$ 表示从 x_k 的任一状态，取决策 u_k 转移到 x_{k+1} 的指标.

基本方程(17－10)给出各后部子过程的最优指标所满足的递推关系，即是说，若已求得系统由第 $k+1$ 段终端某状态转移到终止状态的最优指标 $V_{k+1}(\hat{x}_{k+1})$，便可求得由第 k 段终端某状态转移到终止状态的最优指标 $V_k(x_k)$.

17.3　求解多阶段决策问题的动态规划方法

例 17.3　最短路问题

解　利用基本方程(17－10)逆向递推各阶段的最优决策、最短路及最短路程. 本问题的基本方程为

$$V_k(x_k) = \min_{u_k \in D_k(x_k)} \{L(x_k,u_k) + V_{k+1}(\hat{x}_{k+1})\}, \qquad k = 3,2,1,0,$$

其中 $L(x_k,u_k)$ 是从 x_k 转移至 x_{k+1} 的某状态 $\hat{x}_{k+1}$ 的指标值；

$V_{k+1}(\hat{x}_{k+1})$ 是从 $\hat{x}_{k+1}$ 转移至 x_n 的最优指标值；

$$\hat{x}_{k+1} = T_k(x_k,u_k).$$

1. 解第 3 段($k=3$)

状态：x_3 取 A_3,B_3；

容许决策：$u_3(A_3)=u_3(B_3)=A_4$；

容许决策集合：$D_3(A_3)=D_3(B_3)=\{A_4\}$；

$$V_4(A_4) = 0.$$

1)

$$V_3(A_3) = \min_{u_3\in D_3(A_3)}\{L(A_3,u_3) + V_4(A_4)\}$$

$$= \min_{u_3\in D_3(A_3)} L(A_3,u_3) = L(A_3,A_4) = 30.$$

由此得：$A_3 \to A_4$ 的最短路程为 30；对应的最优决策 $u_3(A_3)=A_4$；最短路为 $A_3\to A_4$.

2)

$$V_3(B_3) = \min_{u_3\in D_3(B_3)} L(B_3,u_3) = L(B_3,A_4) = 40.$$

由此得：$B_3\to A_4$ 的最短路程为 40；对应的最优决策 $u_3(B_3)=A_4$；最短路为 $B_3\to A_4$.

2. 解第 2 段($k=2$)

状态：x_2 取 A_2,B_2,C_2；

容许决策：$u_2(A_2)=u_2(B_2)=u_2(C_2)=A_3,B_3$；

容许决策集合：$\boldsymbol{D}_2(A_2)=\boldsymbol{D}_2(B_2)=\boldsymbol{D}_2(C_2)=\{A_3,B_3\}$；

1)　$$V_2(A_2)=\min_{u_2\in D_2(A_2)}\{L(A_2,u_2)+V_3(\hat{x}_3)\}$$
$$=\min\{L(A_2,A_3)+V_3(A_3),L(A_2,B_2)+V_3(B_3)\}$$
$$=\min\{10+30,40+40\}=40.$$

由此得：$A_2\to A_4$ 的最短路程为 40；对应的最优决策 $u_2(A_2)=A_3$，$u_3(A_3)=A_4$；最短路为 $A_2\to A_3\to A_4$.

2)　$$V_2(B_2)=\min_{u_2\in D_2(B_2)}\{L(B_2,u_2)+V_3(\hat{x}_3)\}$$
$$=\min\{L(B_2,A_3)+V_3(A_3),L(B_2,B_3)+V_3(B_3)\}$$
$$=\min\{60+30,30+40\}=70.$$

由此得：$B_2\to A_4$ 的最短路程为 70；对应的最优决策 $u_3(B_2)=B_3$，$u_2(B_3)=A_4$；最短路为 $B_2\to B_3\to A_4$.

3)　$$V_2(C_2)=\min_{u_2\in D_2(C_2)}\{L(C_2,u_2)+V_3(\hat{x}_3)\}$$
$$=\min\{L(C_2,A_3)+V_3(A_3),L(C_2,B_2)+V_3(B_3)\}$$
$$=\min\{30+30,30+40\}=60.$$

由此得：$C_2\to A_4$ 的最短路程为 60；对应的最优决策 $u_2(C_2)=A_3$，$u_2(A_3)=A_4$；最短路为 $C_2\to A_3\to A_4$.

3. 解第 1 段($k=1$)

状态：x_1 取 A_1,B_1,C_1；

容许决策：$u_1(A_1)=A_2,C_2$，
$u_1(B_1)=u_1(C_1)=A_2,B_2,C_2$；

容许决策集合：$\boldsymbol{D}_1(A_1)=\{A_2,C_2\}$，
$\boldsymbol{D}_1(B_1)=\boldsymbol{D}_1(C_1)=\{A_2,B_2,C_2\}$.

1)　$$V_1(A_1)=\min_{u_1\in D_1(A_1)}\{L(A_1,u_1)+V_2(\hat{x}_2)\}$$
$$=\min\{L(A_1,A_2)+V_2(A_2),L(A_1,C_2)+V_2(C_2)\}$$
$$=\min\{70+40,50+60\}=110.$$

由此得：$A_1\to A_4$ 的最短路程为 110；对应的最优决策 $u_1(A_1)=A_2$，$u_2(A_2)=A_3$，$u_3(A_3)=A_4$ 和 $u_1(A_1)=C_2$，$u_2(C_2)=A_3$，$u_3(A_3)=A_4$；最短路为 $A_1\to A_2\to A_3\to A_4$ 和 $A_1\to C_2\to A_3\to A_4$.

2)　$$V_1(A_1)=\min_{u_1\in D_1(B_1)}\{L(B_1,u_1)+V_2(\hat{x}_2)\}$$
$$=\min\{L(B_1,A_2)+V_2(A_2),L(B_1,B_2)+V_2(B_2),L(B_1,C_2)+V_2(C_2)\}$$
$$=\min\{30+40,20+70,40+60\}=70.$$

由此得：$B_1 \to A_4$ 的最短路程为 70；对应的最优决策 $u_1(B_1) = A_2$，$u_2(A_2) = A_3$，$u_3(A_3) = A_4$；最短路为 $B_1 \to A_2 \to A_3 \to A_4$.

3）类似的讨论可得

$$V_1(C_1) = \min\{L(C_1, u_1) + V_2(\hat{x}_2)\} = 80.$$

最短路为 $C_1 \to A_2 \to A_3 \to A_4$ 和 $C_2 \to B_2 \to b_3 \to A_4$.

4. 解第 0 段（$k=0$）

$$x_0 = A_0, D_0(A_0) = \{A_1, B_1, C_1\},$$

$$\begin{aligned} V_0(x_0) &= \min_{u_0 \in D_0(A_0)} \{L(A_0, u_0) + V_1(\hat{x}_1)\} \\ &= \min\{L(A_0, A_1) + V_1(A_1), L(A_0, B_1) \\ &\quad + V_1(B_1), L(A_0, C_1) + V_1(C_1)\} \\ &= \min\{20 + 110, 40 + 70, 30 + 80\} = 110. \end{aligned}$$

最短路共有 3 条：$A_0 \to B_1 \to A_2 \to A_3 \to A_4$，$A_0 \to C_1 \to A_2 \to A_3 \to A_4$ 和 $A_0 \to C_1 \to B_2 \to B_3 \to B_4$.

在例 17.1 中，为了求出最优路线和最优指标也可以采用“穷举法”，即逐次求出 A_0 到 A_4 的各种可能的路线及其指标值，通过比较指标值确定出最优路线和最优指标. 一般来说，穷举法要比动态规划方法计算量大得多. 以例 17.1 来说，按照穷举法要进行 16 次比较运算和 48 次加法运算，而动态规划方法只需要 10 次比较和 17 次加法运算. 对于阶段数较多，各阶段状态较多的问题，这个差距更为显著. 另一方面，动态规划方法的计算结果较为丰富，不仅得到了全过程从 A_0 至 A_4 的最短路线和最短路程，也得到了各中间状态到 A_4 的最短路线和最短路程，这对于很多实际问题来说往往是很有用的.

例 17.4　多阶段资源分配问题

在前述例 17.2 中，设 $x_0 = 1\ 000$，$g(y) = 5y$，$h(y) = 4y$，$a = 0.8$，$b = 0.9$. 状态方程（17－2）应为

$$x_{k+1} = 0.8u_k + 0.9(x_k - u_k) = 0.9x_k - 0.1u_k. \quad (17-11)$$

解　在本问题，基本方程为

$$\begin{aligned} V_k(x_k) &= \max_{0 \le u_k \le x_k} \{L(x_k, u_k) + V_{k+1}(\hat{x}_{k+1})\} \\ &= \max_{0 \le u_k \le x_k} \{5u_k + 4(x_k - u_k) + V_{k+1}(\hat{x}_{k+1})\}, \end{aligned}$$

$$V_5(x_5) = 0,$$

其中 $\hat{x}_{k+1} = 0.9x_k - 0.1u_k$.

1. $k=4$

$$V_4(x_4) = \max_{0 \le u_4 \le x_4} \{5u_4 + 4(x_4 - u_4)\} = \max_{0 \le u_4 \le x_4} \{u_4 + 4x_4\},$$

因为 u_4+4x_4 是 u_4 的增函数，所以在 $[0,x_4]$ 的右端点取得最大值，即 $V_4(x_4)=5x_4$，相应的决策为 $u_4=x_4$.

2. $k=3$

$$\begin{aligned}V_3(x_3) &= \max_{0\leqslant u_3\leqslant x_3}\{5u_3+4(x_3-u_3)+V_4(\hat{x}_4)\}\\ &= \max_{0\leqslant u_3\leqslant x_3}\{u_3+4x_3+5\hat{x}_4\}\\ &= \max_{0\leqslant u_3\leqslant x_3}\{u_3+4x_3+5(0.9x_3-0.1u_3)\}\\ &= \max_{0\leqslant u_3\leqslant x_3}\{8.5x_3+0.5u_3\} = 9x_3,\end{aligned}$$

相应的决策为 $u_3=x_3$.

3. $k=2$

$$\begin{aligned}V_2(x_2) &= \max_{0\leqslant u_2\leqslant x_2}\{5u_2+4(x_2-u_2)+V_3(\hat{x}_3)\}\\ &= \max_{0\leqslant u_2\leqslant x_2}\{0.1u_2+12.1x_2\} = 12.2x_2,\end{aligned}$$

相应的决策为 $u_2=x_2$.

4. $k=1$

$$\begin{aligned}V_1(x_1) &= \max_{0\leqslant u_1\leqslant x_1}\{5u_1+4(x_1-u_1)+V_2(\hat{x}_2)\}\\ &= \max_{0\leqslant u_1\leqslant x_1}\{-0.22u_1+14.98x_1\}.\end{aligned}$$

因为 $-0.22u_1+14.98x_1$ 是 u_1 的减函数，故在 $[0,x_1]$ 的左端点取得最大值，即 $V_1(x_1)=14.98x_1$，相应的决策为 $u_1=0$.

5. $k=0$，同理得

$$V_0(x_0) = \max_{0\leqslant u_0\leqslant x_0}\{-0.498u_0+17.482x_0\} = 17.482x_0,$$

相应的决策为 $u_0=0$.

因为 $x_0=1\ 000$，故由状态方程(17-11)得

$$\begin{aligned}x_1 &= 0.9x_0-0.1u_0 = 900(\text{台});\\ x_2 &= 0.9x_1-0.1u_1 = 810(\text{台});\\ x_3 &= 0.9x_2-0.1u_2 = 648(\text{台});\\ x_4 &= 0.9x_3-0.1u_3 = 518.4(\text{台});\\ x_5 &= 0.9x_4-0.1u_4 = 414.72(\text{台})\end{aligned}$$

（注：$x_4=518.4$ 台中的 0.4 应理解为有一台机器只能使用 0.4 年即将报

废，同理 414.72 台也应如此理解）.

结论：各年安排生产两种产品的机器台数如表 17－1 所示：

表 17－1

年度	1	2	3	4	5
生产 A 种产品机器数	$u_0=0$	$u_1=0$	$u_2=810$	$u_3=648$	$u_4=518.4$
生产 B 种产品机器数	$x_0-u_0=1\ 000$	$x_1-u_1=900$	$x_2-u_2=0$	$x_3-u_3=0$	$x_4-u_4=0$

五年的最高收入为 $V_0(x_0)=17.482x_0=17\ 482$（万元）$=1.748\ 2$ 亿元.

最后我们将求解例 17.2 的计算过程归纳为如下步骤：

1. 由动态规划基本方程

$$\begin{cases} V_k(x_k)=\max\{5u_k+4(x_k-u_k)+V_{k+1}(x_{k+1})\}, \\ V_5(x_5)=0. \end{cases}$$

求得多阶段决策问题中决策变量的形式解：

$$u_4^*=x_4,\ u_3^*=x_3,\ u_2^*=x_2,\ u_1^*=0,\ u_0^*=0.$$

2. 将此形式解代入状态方程得 u^* 的相应状态：

$x_1^*=900,\ x_2^*=810,\ x_3^*=648,\ x_4^*=518.4,\ x_5^*=414.72,$

3. 再将 x^* 代入形式解得问题的反馈解：

$$u_0^*=0, u_1^*=0, u_2^*=810, u_3^*=648, u_4^*=518.4.$$

17.4 多阶段决策问题实例分析

在企业管理中，生产库存是一个普遍存在的问题. 若某产品库存量过小，当销售量大幅度增加时，将形成脱销，使企业失去了多获得利润的机会；若库存量过大，除了需付出保管费用外，还要积压资金，不利于企业的发展，因此库存管理对企业和国民经济都具有重大意义. 库存量的变化除受销售量的影响外，还与产品生产紧密相连，因此必须综合地研究库存系统的管理问题. 下面列举离散系统库存的数学模型，它们是多阶段决策系统的典型问题.

例 17.5 生产—库存管理系统

设 $u(i)$ 表示某厂第 i 季度（或月）的生产产品的数量；

$x(i)$ 表示该厂第 i 季度（或月）的库存量；

$S(i)$表示该厂第 i 季度(或月)的销售量(已知).

显然,控制变量 $u(i)$,状态变量 $x(i)$与外生变量 $S(i)$之间应满足如下状态方程:

$$x(i+1) = x(i) + u(i) - S(i) \quad (i = \overline{1,4} \text{ 或 } i = \overline{1,12}). \tag{17-12}$$

现在的问题是:在满足状态方程(17-12)的条件下,求每个季度(或月)的最优生产量 $u^*(i)$和相应的最优库存量 $x^*(i)$,使得生产费用与库存量用之和为最小. 由给定的比例系数 α,β 和阶数 k_1,k_2,求 $u^*(i)$和 $x^*(i)$使得

$$J = \sum_{i=1}^{4\text{或}12} (\alpha u(i)^{k_1} + \beta x(i)^{k_2}) = \min. \tag{17-13}$$

若生产量和库存量都有限,还应加上约束条件:

$$0 \leqslant u(i) \leqslant N,\ 0 \leqslant x(i) \leqslant M,\ i = \overline{1,4} \text{ 或 } i = \overline{1,12}.$$

例 17.6 库存—销售系统

若某商店,货的库存最多可有 $\overline{M}$ 单位商品,已知有存货 M 单位,并预知该商品的进价和售价.

设 $x(i)$为第 i 月存货量;$u(i)$为第 i 月售出量;$v(i)$为第 i 月进货量,则状态变量 $x(i)$与控制变量 $u(i)$,$v(i)$之间满足以下关系:

$$x(i+1) = x(i) + v(i) - u(i), \tag{17-14}$$

约束条件: $0 \leqslant u(i) \leqslant v(x)$,

$$0 \leqslant u(i) \leqslant M - (x(i) - u(i)),\quad k = 1,\cdots,N. \tag{17-15}$$

现在的问题是:满足状态方程(17-14)及约束条件(17-15)的情形下,求最优策略 $u^*(i)$和 $v^*(i)$使商品获得最大利润,即使指标函数

$$J = \sum_{k=1}^{N} [(p_k u(i) - q_k v(i))] = \max, \tag{17-16}$$

其中 p_k 为单位商品的售出价,q_k 为单位商品的购入价. 该问题中,每阶段有两个决策变量 $u(i)$,$v(i)$.

例 17.7 购销——生产——库存系统

当生产企业面对原材料(零配件)的购进、生产或加工,原材料(零配件)及企业产品的库存,产品销售等一系列问题时,则整个系统必须分解为两个子系统来处理:

1. 购存产子系统:该子系统实施以产订购的最优策略,即在确定生产量的前提下实施对原材料的订购和库存的管理.

设 $u(i)$表示第 i 时刻的进货量;$x(i)$表示第 i 时刻的库存量;$S(i)$表示第 i 时刻的生产量,状态变量 $x(i)$,控制变量 $u(i)$及已知外生变量

$S(i)$满足状态方程

$$x(i+1) = x(i) + u(i) - S(i) \quad (\text{外生变量 } S(i) \text{ 为预测值}). \tag{17-17}$$

现在的问题是:确定最优订货批量$u^*(i)$及最优库存量$x^*(i)$,使进货量$u(i)$趋于$u^*(i)$,实际库存$x(i)$趋于$x^*(i)$.即使指标函数

$$J = \sum_{i=0}^{N-1} [\alpha(x(i) - x^*(i))^2 + \beta(u(i) - u^*(i))^2] = \min, \tag{17-18}$$

其中,α,β为加权系数.

当原材料(零配件),产品为多种时,该方程的变量为向量,α、β亦为向量.

2. 产存销子系统:该子系统实施以销定产的最优策略,即在已知销售量的前提下实施对生产产量和库存量的管理.

设$u(i)$表示第i时刻的生产量;$x(i)$表示第i时刻的库存量;$S(i)$表示第i时刻的销售量.

控制变量$u(i)$,状态变量$x(i)$及已知外生变量$S(i)$满足状态方程:

$$x(i+1) = x(i) + u(i) - S(i) \quad (\text{外生变量 } S(i) \text{ 为预测值}). \tag{17-19}$$

现在的问题是:确定最优订货批量$u^*(i)$及最优库存量$x^*(i)$,使进货量$u(i)$趋于$u^*(i)$,实际库存$x(i)$趋于$x^*(i)$.即使

$$J = \sum_{i=0}^{N-1} [p(x(i) - x^*(i))^2 + q(u(i) - u^*(i))^2] = \min, \tag{17-20}$$

其中,p,q为加权系数.

3. 两个子系统之间的关系

第2子系统得到的最优控制$u^*(i)$是确定第1子系统的外生变量$S(i)$的依据.求解顺序应是先由销售额$S(i)$的预测值求出第2子系统的最优生产量$u^*(i)$,以此定出第1子系统的生产量$S(i)$,再求出第1子系统的最优购货量$u^*(i)$.

下面我们利用动态规划的基本方程(17-10)给出例17.3的求解步骤.

解 1. 确定模型中的参数α,β,k_1,k_2.

在(17-1?)中,α,k_1与生产成本有关;β,k_2与库存成本有关.这些参数通常用相关的经验公式给出,在本例中取$k_1=1,k_2=1,\alpha,\beta$为给定常数.

2. 销售量 $S(i)$ $(i=1,\cdots,N)$ 的预测.

设将一年分成 N 个时段,若按月份考虑 $N=12$,按季节 $N=4$,通过搜集厂销售产品的历史数据,用拟合、回归方法(中长期预测)或用时间序列分析法(短期预测)可以预测出模型中所需要的销售量 $S(i)$.

3. 库存初值 $x(1)$ 及终值 $x(N+1)$ 的确定.

第一时段初产品的库存认为是已知的,即初值 $x(1)=x_0$,现在要研究终值 $x(N+1)$ 的确定,即第 N 时段末的库存.

设每一时段末的库存等于下一时段初的库存,即

$$x(i)_{末} = x(i+1)_{初},$$

故有

$$x(N+1)_{初} = x(N)_{末}.$$

$x(N)_{末}$ 与预定货水平有关,我们把 $x(N+1)$ 认为是 $x(N)$ 最后一次订货的最佳预定水准 A,即令 $x(N+1)=A$.

4. 按基本方程求决策变量 $u(i)$ 的最优解的形式解 $u^*(i)$ 和相应的状态 $x^*(i)$,进而确定形式解的反馈解.

具体计算过程归纳成以下算法.

算法 17.1

① 准备初值 $x(1)=x_0, x(N+1)=A, S(i)=S_i, i=1,2,\cdots,N$.

② 计算 $u^*(i)$ 的形式解 $(i=N,N-1,\cdots,1)$

(1) 解第 N 段的基本方程 $V_1^*(x(N))=\min\{\alpha u^2(N)+\beta x(N)\}$,

$$(17-21)$$

即求 $u^*(N)$ 的形式解以及对应的最优函数 $V_1^*(x(N))$.

由第 N 段的状态方程 $x(N+1)=x(N)+u(N)-S_N$,以及库存终值条件 $x(N+1)=A$ 得 $x(N)+u(N)-S_N=A$.

满足此方程的解 $u(N)$ 只有一解,亦即最优决策 $u^*(N)=S_N-x(N)+A$.

将 $u^*(N)$ 代入(17-21)得

$$V_1^*(x(N)) = \alpha[S_N - x(N) + A]^2 + \beta x(N). \qquad (17-22)$$

(2) 解第 $N-1$ 段的基本方程

$$V_2^*(x(N-1)) = \min\{\alpha u^2(N-1) + \beta x(N-1) + V_1^*(x(N))\}, \qquad (17-23)$$

即求 $u^*(N-1)$ 的形式解以及对应的最优函数 $V_2^*(x(N-1))$.

据第 $N-1$ 段的状态方程　$x(N)=x(N-1)+u(N-1)-S_{N-1}$,

将 $x(N)$ 代入(17-22),再将 $V_1^*(x(N))$ 代入(17-23),则 $V_2^*(x(N-1))$ 的表达式中只含 $u(N-1)$, $x(N-1)$ 以及常数 α,β,S_{N-1},S_N,再由

$\frac{\partial\{\cdot\}}{\partial u(N-1)}=0$，求解 $u^*(N-1)$ 的表达式，它是 $x(N-1)$ 的多项式，进而 $V_2^*(x(N-1))$ 也是 $x(N-1)$ 的多项式.

…………

进行至第 $(N-1)$ 步，已求得 $u^*(2)$ 的形式解（它是 $x(2)$ 的多项式）以及对应的最优函数 $V_{N-1}(x(2))$（它也是 $x(2)$ 的多项式）.

(N) 解第 1 段的基本方程

$$V_N^*(x(1)) = \min\{\alpha u^2(1) + \beta x(1) + V_{N-1}^*(x(2))\}, \tag{17-24}$$

即求 $u^*(1)$ 的形式解以及对应的最优函数 $V_N^*(x(1))$.

据第 1 段的状态方程 $x(2)=x(1)+u(1)-S_1$，将 $x(2)$ 代入 $V_{N-1}^*(x(2))$ 中，再将 $V_{N-1}^*(x(2))$ 的表达式代入式(17－24)，再由 $\frac{\partial\{\cdot\}}{\partial u(1)}=0$ 解得 $u^*(1)$ 的表达式，它是 $x(1)$ 的多项式，进而 $V_N^*(x(1))$ 亦是 $x(1)$ 的多项式（因为 $x(1)=x=0$ 为已知，故 $u^*(1)$ 为可知）.

③ 计算相应的状态 $x^*(i)$ 以及 $u^*(i)$ $(i=1,2,\cdots,N)$ 的反馈解.

(1) $x^*(1)=x(1)=x_0$，由 2 的第 N 步得到 $u^*(1)$ 的反馈解；

(2) 列出第 1 时段的状态方程

$$x(2) = x(1) + u(1) - S(1),$$

将 $x^*(1)$，$u^*(1)$ 代入得 $x^*(2)$；再代入 2 的第 $N-1$ 步得到 $u^*(2)$ 的反馈解；

(3) 列出第 2 时段的状态方程

$$x(3) = x(2) + N(2) - S(2),$$

将 $x^*(2)$，$u^*(2)$ 代入得 $x^*(3)$；再代入 2 的第 $N-2$ 步得到 $u^*(3)$ 的反馈解；

…………

(N) 列出第 $N-1$ 时段的状态方程，

$$x(N) = x(N-1) + u(N-1) - S(N-1),$$

将 $x^*(N-1)$，$u^*(N-1)$ 代入得 $x^*(N)$；再代入②的第 1 步得到 $u^*(N)$ 的反馈解.

17.5 离散线性二次型系统的动态规划方法

设线性定常系统的状态方程为

$$\begin{cases} x(k+1) = Ax(k) + Bu(k), \\ x(0) = x^{(0)}, \quad k = 0,1,\cdots,n-1. \end{cases} \tag{17-25}$$

指标函数为

$$J = x(n)^{\mathrm{T}} S_0 x(n) + \sum_{k=0}^{n-1} [x(k)^{\mathrm{T}} Q x(k) + u(k)^{\mathrm{T}} R u(k)], \tag{17-26}$$

其中 A 为 $n \times n$ 阶常数矩阵，B 为 $n \times r$ 阶常数矩阵. S_0 和 Q 为 $n \times n$ 阶半正定常数矩阵，R 为 $r \times r$ 阶正定常数矩阵.

状态变量 $x(k)$ 为 n 维向量，决策向量 $u(k)$ 为 r 维向量.

现在的问题是：求最优决策 $u^*(k)$ 及相应的状态 $x^*(k)$，使得目标 J 达到最小.

解　系统(17－25)、(17－26)的动态规划基本方程为：

$$\begin{cases} V(x(k)) = \min\{x(k)^{\mathrm{T}} Q x(k) + u(k)^{\mathrm{T}} R u(k) + V(x(k+1))\}, \\ V(x(n)) = x(n)^{\mathrm{T}} S_0 x(n), \quad k = n-1, n-2, \cdots, 0. \end{cases} \tag{17-27}$$

其中 $S_0 = S(n)$（给定），$x(k+1)$ 据(17－25)，由 $x(k)$，$u(k)$ 确定.

根据状态方程(17－25)和基本方程(17－27)可以推导出如下递推式（推导过程略）：

$$\begin{cases} L(k) = [B^{\mathrm{T}} S(k+1) B + R]^{-1} B^{\mathrm{T}} S(k+1) A, & (17-28) \\ S(k) = [A - BL(k)]^{\mathrm{T}} S(k+1) [A - BL(k)] \\ \qquad\quad + L(k)^{\mathrm{T}} R L(k) + Q, & (17-29) \\ u^*(k) = -L(k) x(k), & (17-30) \\ V(x(k)) = x(k)^{\mathrm{T}} S(k) x(k). & (17-31) \end{cases}$$

据(17－28)，(17－29)逆向递推可得：

$$L(n-1),\ L(n-2),\ \cdots,\ L(1),\ L(0),$$
$$S(n-1),\ S(n-2),\ \cdots,\ S(1),\ S(0).$$

递推过程按并行方案进行：

$$S(n) = S_0 \rightarrow L(n-1) \rightarrow S(n-1) \rightarrow L(n-2) \rightarrow S(n-2) \rightarrow \cdots \rightarrow L(0) \rightarrow S(0),$$

求出 $L(k)$ $(k=0,1,\cdots,n-1)$ 后按如下联立公式计算 $u^*(k)$ $(k=0,1,\cdots,n-1)$，以及 $x^*(k)$ $(k=0,1,\cdots,n)$

$$\begin{cases} x(k+1) = Ax(k) + Bu(k), & (17-25) \\ u^*(k) = -L(k) x(k), & (17-30) \\ k = 0,1,\cdots,n-1, \end{cases}$$

其中 $L(k)$ $(k=0,1,\cdots,n-1)$ 为给定值. 上述计算过程归纳成为如下算法:

算法 17.2

① 给定初值 $x(0)=x^{(0)}$, $S(n)=S_0(\neq S(0))$ 及 $A,B,Q,R,x(n)$;

② 令 $k=n-1,n-2,\cdots,0$ 计算

$$\begin{cases}L(k)=[B^{T}S(k+1)B+R]^{-1}B^{T}S(k+1)A,\\ S(k)=[A-BL(k)]^{T}S(k+1)[A-BL(k)]+L^{T}(k)RL(k)+Q,\end{cases}$$

得到 $L(k)$ $(k=0,1,\cdots,n-1)$;

③ 令 $k=0,1,\cdots,n-1$ 计算 $u^*(k)$ 的形式解

$$u^*(k)=-L(k)x(k),$$

其中 $x(0)$ 为已知, $x(k)$ $(k=1,\cdots,n-1)$ 为未知;

④ 给定 $x(n)$, 将形式解 $u^*(k)$ 代入状态方程 (17-25), 求 $x^*(k)$ $(k=n-1,n-2,\cdots,1)$,

计算过程图示如下:

$$x(n)\xrightarrow[\text{代入}(17-25)]{u^*(n-1)=-L(n-1)x(n-1)}x^*(n-1)\xrightarrow[\text{代入}(17-25)]{u^*(n-2)=-L(n-2)x(n-2)}$$

$$x^*(n-2)\rightarrow\cdots\rightarrow\xrightarrow[\text{代入}(17-25)]{u^*(1)=-L(1)x(1)}x^*(1);$$

⑤ 将 $x^*(k)$ $(k=0,1,\cdots,n-1)$ 代入形式解 $u^*(k)=-L(k)x^*(k)$, 求出反馈解 $u^*(k)$ $(k=0,1,\cdots,n-1)$;

⑥ 据 (17-31) 计算最优目标值

$$J^*=V(x^*(0))=x^*(0)^{T}S(0)x^*(0).$$

例 17.8　已知线性定常系统的状态方程为

$$\begin{cases}x(k+1)=ax(k)+bu(k),\\ x(0)=x_0,\quad k=0,1,2,\end{cases}$$

性能指标为 $J=\sum_{k=0}^{2}[x^2(k)+cu^2(k)]$, 求使 J 取最小值的最优决策 $u^*(k)$ $(k=0,1,2)$.

解　(1) 给定初值 $x(0)=x^{(0)}$, $S(3)=S_0=0$ 及 $A=a,B=b,Q=I$, $R=c$;

(2) 令 $k=2,1,0$ 计算

$$\begin{cases}L(k)=[b^2S(k+1)+c]^{-1}bS(k+1)a,\\ S(k)=[a-bL(k)]^2S(k+1)+cL^2(k)+1,\\ S(3)=0.\end{cases}$$

$k=2$, 得 $L(2)=0$, $S(2)=1$,

$k=1$,得 $L(1)=\dfrac{ab}{b^2+c}$,$S(1)=\dfrac{a^2c}{b^2+c}+1$,

$k=0$,得 $L(0)=\left[b^2\left(\dfrac{a^2c}{b^2+c}+1\right)+c\right]^{-1}b\left(\dfrac{a^2c}{b^2+c}+1\right)a$

$$=\frac{ab(a^2c+b^2+c)}{b^2+a^2b^2c+2b^2c+c^2},$$

$$S(0)=[a-bL(0)]^2\left(\frac{a^2c}{b+c}+1\right)+cL^2(0)+1,$$

(3) 令 $k=0,1,2$ 计算 $u^*(k)$的形式解

$$u^*(0)=-L(0)x(0)=-\frac{ab(a^2c+b^2+c)}{b^2+a^2b^2c+2b^2c+c^2}x_0,$$

$$u^*(1)=-L(1)x(1)=-\frac{ab}{b^2+c}x(1),$$

$$u^*(2)=-L(2)x(2)=0.$$

(4) 将 $u^*(k)$ $(k=0,1,2)$代入状态方程得

$$x^*(2)=\frac{1}{a}x(3),$$

$$x^*(1)=\frac{b^2+c}{a^2(b^2+c-b)}x(3),$$

$$x^*(0)=\frac{b^2+c}{a^3(b^2+c-b)}+\frac{b^2(a^2c+b^2+c)}{b^2+a^2b^2c+2b^2+c^2}x(0).$$

(5) 将 $x^*(k)$ $(k=0,1,2)$代入形式解得 $u^*(k)$的反馈解:

$$\begin{cases} u^*(0)=-\dfrac{ab(a^2c+b^2+c)}{b^2+a^2b^2c+2b^2c+c^2}x^*(0), \\ u^*(1)=-\dfrac{ab}{b^2+c}x^*(1), \\ u^*(2)=-L(2)x^*(2)=0. \end{cases}$$

(6) 计算最优目标值

$J^*=V(x^*(0))=x^*(0)^{\mathrm{T}}S(0)x^*(0)$,其中 $S(0)$在(2)中求得,$x^*(0)$在(4)中求得.

第 18 章　现代优化方法简介

本章介绍求解复杂优化问题的算法，与前面各章介绍的常用优化算法相比较，它们具有如下的特点：第一，对目标函数和约束函数不必附加可解析性的条件，对于目标函数而言，甚至不要求具有显式表达式，只需要在所计算的点处提供函数值；第二，对于约束变量可以取离散值，比如整数值，或取某些特殊值 0 或 1；第三，在通常情况下，这些算法能够求解全局最优点.

18.1　随机试验法

随机试验法是用随机方法生产试验点，再从试验点中选出满足约束条件的点，进而求出最优点的一种方法. 具体方法是：

设问题为

$$\min f(\boldsymbol{x}),$$
$$\text{s.t. } g_i(\boldsymbol{x}) \leqslant 0, i = 1,2,\cdots,m$$

则首先用随机方法生产试验点 $\{\boldsymbol{x}_k\}$，然后从中找出满足约束条件

$$\text{s.t. } g_i(\boldsymbol{x}) \leqslant 0, i = 1,2,\cdots,m$$

的点 $\{\boldsymbol{x}_{k_p}\}$，并求出使得

$$\min_{k_p}\{f(\boldsymbol{x}_{k_p})\} = f(\boldsymbol{x}^*)$$

成立的最优解 $\boldsymbol{x}^*$. 但这种方法因选点过程计算量较大，所以当问题的维数很大时计算量是非常大的.

随机试验法的算法很多，但这里仅介绍三种，按这三种算法依次进行下去便可以得到较准确的最优解. 但这些算法必须满足两个条件：

(1) 给出最优解 $\boldsymbol{x}^*$ 的估计范围，即给出 $a_j, b_j (j = 1,2,\cdots,m)$ 使

$$a_j < x_j^* < b_j (j = 1,2,\cdots,m);$$

(2) 给出产生(0,1)区间内均匀分布的伪随机数的方法. 通常一般计算机的软件中都给有产生(0,1)区间内均匀分布的伪随机数的程序.

这里仅介绍一种求伪随机数的方法——“乘同余法”.

(1) 用递推公式产生数列 t_i：

$$t_i = A \cdot t_{i-1} (\text{mod } m),$$

其中，t_i 表示正整数 $A \cdot t_{i-1}$ 被整数 m 去除所剩下的余数，此式称为按模为 m 的同余式．且初值 t_0，乘子 A 和模 m 都取非负的整数．

(2) 用 m 除 t_i，即得区间(0,1)上的伪随机数

$$r_i = \frac{t_i}{m}.$$

例如，若取 $t_0 = 74, A = 13, m = 75$，则可算得：

$$t_1 = 13 \times 74/75 \text{ 的余数} = 62,$$
$$t_2 = 13 \times 62/75 \text{ 的余数} = 56,$$
$$t_3 = 13 \times 56/75 \text{ 的余数} = 53,$$
$$\cdots\cdots\cdots\cdots$$
$$r_1 = t_1/m = 62/75 = 0.8267,$$
$$r_2 = t_2/m = 56/75 = 0.7467,$$
$$r_3 = t_3/m = 53/75 = 0.7067,$$
$$\cdots\cdots\cdots\cdots$$

假定有了 n 个(0,1)区间内均匀分布的伪随机数 $\lambda_j (j = 1,2,\cdots,n)$，便可得到一个点

$$\bar{x} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n)^{\mathrm{T}},$$

其中 $\bar{x}_j = a_j + \lambda_j(b_j - a_j)$，

若 $$g_i(\bar{x}) \leqslant 0 (i = 1,2,\cdots,m),$$

则 $\bar{x}$ 便是可行点．如此便可用以上随机方法找出点列 $\{x_k\}$，再在点列中找出可行点列 $\{x_{k_p}\}$，记 $\min\limits_{k_p}\{f(x_{k_p})\} = f(x^*)$，当 k_p 趋于无穷大时，视 x^* 为所求的极小值点．

分析以上过程知，每产生一点 x_k，要用到 n 个随机数，运算量较大，为减少运算量，使得每产生一点只用一个新的随机数，令

$$x_k = (x_1^k, x_2^k, \cdots, x_n^k)^{\mathrm{T}},$$
$$x_{k+1} = (x_1^{k+1}, x_2^{k+1}, \cdots, x_n^{k+1})^{\mathrm{T}},$$

分量关系满足

$$x_j^k = a_j + \lambda_j(b_j - a_j), \quad j = 1,2,\cdots,n,$$
$$x_j^{k+1} = a_j + \lambda_{j+1}(b_j - a_j), \quad j = 1,2,\cdots,n-1,$$
$$x_n^{k+1} = a_n + \lambda(b_n - a_n),$$

其中，λ 为新生成的(0,1)区间内均匀分布的伪随机数 λ．

算法 18.1

算法中假定 k 为随机点的总数；k_p 为实验的可行点总数；k_m 为最小

解未作改变的次数；M 为给定的正数；N 为给定的正数，最小解如 N 次不变，则认为求出了最优解；N_1 为给定的正数，如产生 N_1 个随机点仍无可行点，则认为问题无可行解，控制输出信息.

① $0 \Rightarrow k \Rightarrow k_p$；$M \Rightarrow p$，生成 $n-1$ 个 $(0,1)$ 区间内均匀分布的伪随机数 $\lambda_j (j=2,\cdots,n)$；

② $\lambda_j \Rightarrow \lambda_{j-1} (j=2,3,\cdots,n)$，产生一个 $(0,1)$ 区间内均匀分布的伪随机数 λ，$\lambda \Rightarrow \lambda_n$，求 $x_j = a_j + \lambda_j (b_j - a_j)$，$j=1,2,\cdots,n$

记 $$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}.$$

③ 若 $g_i(\boldsymbol{x}) \leqslant 0 (i=1,2,3,\cdots,m)$？

是，转④；

否，转⑥；

④ $k_p + 1 \Rightarrow k_p$，求 $f(\boldsymbol{x})$. 当 $f(\boldsymbol{x}) < p$ 时，$\boldsymbol{x} \Rightarrow \boldsymbol{x}^*$，$f(\boldsymbol{x}) \Rightarrow p$，$0 \Rightarrow k_m$；当 $f(\boldsymbol{x}) \geqslant p$ 时，$k_m + 1 \Rightarrow k_m$；

⑤ 若 $k_m \geqslant N$，认为求出了最优解，停止计算；若 $k_m < N$，转⑥；

⑥ $k+1 \Rightarrow k$，当 $k \geqslant N_1$ 时，(1) 若 $k_p = 0$，认为问题无可行解，停止；(2) 若 $k_p > 0$，则 $2N_1 \Rightarrow N_1$，转②；当 $k < N_1$，转②.

在算法 18.1 中，$a_j, b_j (j=1,2,\cdots,n)$ 不变，这不利于对 $\boldsymbol{x}^*$ 进一步估计，且经过计算后会逐渐对 $\boldsymbol{x}^*$ 有一个更准确的估计，此时的区间运算值 $a_j, b_j (j=1,2,\cdots,n)$ 也应逐步靠近，若使算法 18.1 中计算到第 N_1 步的结果作为 $\boldsymbol{x}^*$ 的新的估计值 $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n)^{\mathrm{T}}$，且对每一个变量估计一个区间长度 l_j，令

$$a_j = \bar{x}_j - \frac{l_j}{2}, b_j = \bar{x}_j + \frac{l_j}{2}, (j = 1,2,\cdots,n),$$

得算法 18.2.

算法 18.2

① $0 \Rightarrow k \Rightarrow k_p$；$N_1 \Rightarrow k_r$，生成 $n-1$ 个 $(0,1)$ 区间内均匀分布的伪随机数 $\lambda_j (j=2,\cdots,n)$，$M \Rightarrow p$；

② 对给定的 $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n)^{\mathrm{T}}$，有 $\bar{x}_j - \frac{l_j}{2} \Rightarrow a_j$，$\bar{x}_j + \frac{l_j}{2} \Rightarrow b_j$，$(j=1,2,\cdots,n)$

③ $\lambda_j \Rightarrow \lambda_{j-1} (j=2,3,\cdots,n)$，产生一个 $(0,1)$ 区间内均匀分布的伪随机数 λ，$\lambda \Rightarrow \lambda_n$，有 $x_j = a_j + \lambda_j (b_j - a_j)$，$j=1,2,\cdots,n$.

记 $$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}},$$

④ 若 $g_i(\boldsymbol{x}) \leqslant 0 (i=1,2,3,\cdots,m)$？

是，转⑤；

否，若有 i_0 使 $g_{i_0}(\boldsymbol{x})>0$，转⑦；

⑤ $k_p+1\Rightarrow k_p$，求 $f(\boldsymbol{x})$. 当 $f(\boldsymbol{x})<p$ 时，$\boldsymbol{x}\Rightarrow\boldsymbol{x}^*$，$f(\boldsymbol{x})\Rightarrow p$，$0\Rightarrow k_m$；当 $f(\boldsymbol{x})\geqslant p$ 时，$k_m+1\Rightarrow k_m$；

⑥ 若 $k_m\geqslant N$，认为求出了最优解，停止计算；若 $k_m<N$，转⑦；

⑦ $k+1\Rightarrow k$，当 $k\geqslant k_v$ 时，(1) 若 $k_p=0$，认为问题无可行解，停止；(2) 若 $k_p>0$，则 $k_v+N_1\Rightarrow k_v$，$\boldsymbol{x}^*\Rightarrow\bar{\boldsymbol{x}}$，$\alpha l_j\Rightarrow l_j(j=1,2,\cdots,n)$，转②；当 $k<k_v$，转③.

注　算法 18.2 中 α 为一个小于 1 的正数；k 存放控制改变区间长 l_j 的迭代次数.

为了更好的利用所找出的 $\boldsymbol{x}^*$，又提出了随机方向法. 把前 N_1 次找出的最优点 $\boldsymbol{x}$ 与再作 N_1 次后求出的 $\boldsymbol{x}^*$ 比较，若比较 $\|\bar{\boldsymbol{x}}-\boldsymbol{x}^*\|>\varepsilon$，则沿方向 $\boldsymbol{p}=\boldsymbol{x}^*-\bar{\boldsymbol{x}}$ 从 $\boldsymbol{x}^*$ 出发作一维搜索求出 $\boldsymbol{x}^0$，用 $\boldsymbol{x}^0$ 代替 $\bar{\boldsymbol{x}}$ 再计算下去. 这样做可以加速收敛速度. 相应地得算法 18.3.

算法 18.3

① $0\Rightarrow k\Rightarrow k_p$；$N_1\Rightarrow k_r$，生成 $n-1$ 个 (0,1) 区间内均匀分布的伪随机数 $\lambda_j(j=2,\cdots,n)$，$M\Rightarrow p$；

② 对给定的 $\bar{\boldsymbol{x}}=(\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n)^{\mathrm{T}}$，求 $\bar{x}_j-\dfrac{l_j}{2}\Rightarrow a_j$，$\bar{x}_j+\dfrac{l_j}{2}\Rightarrow b_j(j=1,2,\cdots,n)$；

③ $\lambda_j\Rightarrow\lambda_{j-1}(j=2,3,\cdots,n)$，产生一个 (0,1) 区间内均匀分布的伪随机数 λ，$\lambda\Rightarrow\lambda_n$，

求
$$x_j=a_j+\lambda_j(b_j-a_j),\ j=1,2,\cdots,n.$$
记
$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}},$$

④ 若 $g_i(x)\leqslant 0(i=1,2,3,\cdots,m)$，转⑤；否则，若有 i_0 使 $g_{i_0}(x)>0$，转⑦；

⑤ $k_p+1\Rightarrow k_p$；求 $f(\boldsymbol{x})$. 当 $f(\boldsymbol{x})<\boldsymbol{p}$ 时，$\boldsymbol{x}\Rightarrow\boldsymbol{x}^*$，$f(\boldsymbol{x})\Rightarrow\boldsymbol{p}$，$0\Rightarrow k_m$；当 $f(\boldsymbol{x})\geqslant\boldsymbol{p}$ 时，$k_m+1\Rightarrow k_m$；

⑥ 若 $k_m\geqslant N$，认为求出了最优解，停止计算；若 $k_m<N$，转⑦；

⑦ $k+1\Rightarrow k$，当 $k\geqslant k_v$ 时，(1) 若 $k_p=0$，认为问题无可行解，停止；(2) 若 $k_p>0$，则比较 $\|\bar{\boldsymbol{x}}-\boldsymbol{x}^*\|$ 与给定的 ε，若 $\|\bar{\boldsymbol{x}}-\boldsymbol{x}^*\|>\varepsilon$，转⑧；若 $\|\bar{\boldsymbol{x}}-\boldsymbol{x}^*\|\leqslant\varepsilon$，则 $\boldsymbol{x}^*\Rightarrow\bar{\boldsymbol{x}}$，$k_v+N_1\Rightarrow k_v$，$\alpha l_j\Rightarrow\alpha l_j(j=1,2,\cdots,n)$ 转②；当 $k<k_v$，转③；

⑧ 求 $\boldsymbol{p}=\boldsymbol{x}^*-\bar{\boldsymbol{x}}$，即可行域为 D，用一维搜索求

$f(\boldsymbol{x}^0)\Rightarrow\min\{f(\boldsymbol{x}^*+\alpha\boldsymbol{p})\mid\boldsymbol{x}^*+\alpha\boldsymbol{p}\in D\}$，

$f(\boldsymbol{x}^0)\Rightarrow\boldsymbol{p}$，$\boldsymbol{x}^0\Rightarrow\bar{\boldsymbol{x}}$，$\boldsymbol{x}^0\Rightarrow\boldsymbol{x}^*$，$k_v+N_1\Rightarrow k_v$，$\alpha l_j\Rightarrow l_j(j=1,2,\cdots,n)$，转②.

18.2 禁忌搜索算法

禁忌搜索算法(简称 TS)是在 1986 年由 Glover 和 Hansen 首次提出的求最优解的算法. 它是一种人工智能算法,是局部邻域搜索算法的推广,是人工智能在组合优化算法中的一个成功应用.

禁忌搜索算法的一个主要思想是标记已得到的局部最优解,并在进一步的迭代中避开这些局部最优解. 所谓的禁忌,就是禁止重复前面的工作. 为了避开局部邻域搜索陷入局部最优,禁忌搜索算法设计了一种禁忌表,记录已经到达过的局部最优点. 在下一次的搜索中,就利用禁忌表中的信息,不再或者有选择地搜索这些点,以此跳出局部最优点.

18.2.1 禁忌搜索算法的主要步骤

禁忌搜索算法是从一个初始可行解出发,选择一系列特定的搜索方向(移动)作为试探,选择实现让特定的目标函数值减少最多的移动,一次移动产生一个实验解. 为了避免陷入局部最优解,搜索中采用对已经进行的优化过程进行记录和选择,指导下一步的搜索方向,这就是建立的 Tabu 表. Tabu 表中保存了最近若干次迭代过程中所实现的移动的反方向移动,凡是在 Tabu 表中移动,在当前迭代过程中则不能实现. 这样就防止了重复搜索,从而避免了算法陷入局部最优解. 同时,为了尽量不错过产生最优解的移动,算法采用了特赦原则.

假设算法是用以解决如下组合优化问题:

$$\min f(x),$$
$$\text{s.t.} \begin{cases} g(x) \geqslant 0, \\ x \in D, \end{cases}$$

则禁忌搜索算法的程序框图如图 18-1

禁忌搜索算法的主要步骤为:

算法 18.4

① 选定一个初始解 x_{best},令禁忌表 $T=\phi$;

② 终止准则?

是,转④;

否,在一定的搜索方向产生移动值 f_{move},在 x_{best} 的邻域 $D(x_{best})$ 中选择满足禁忌要求的候选集 $CD(x_{best})$,转③;

③ 在 $CD(x_{best})$ 中选一个评价值最好的解 x_{next},令 $x_{best}=x_{next}$,更新禁忌表 T,转②;

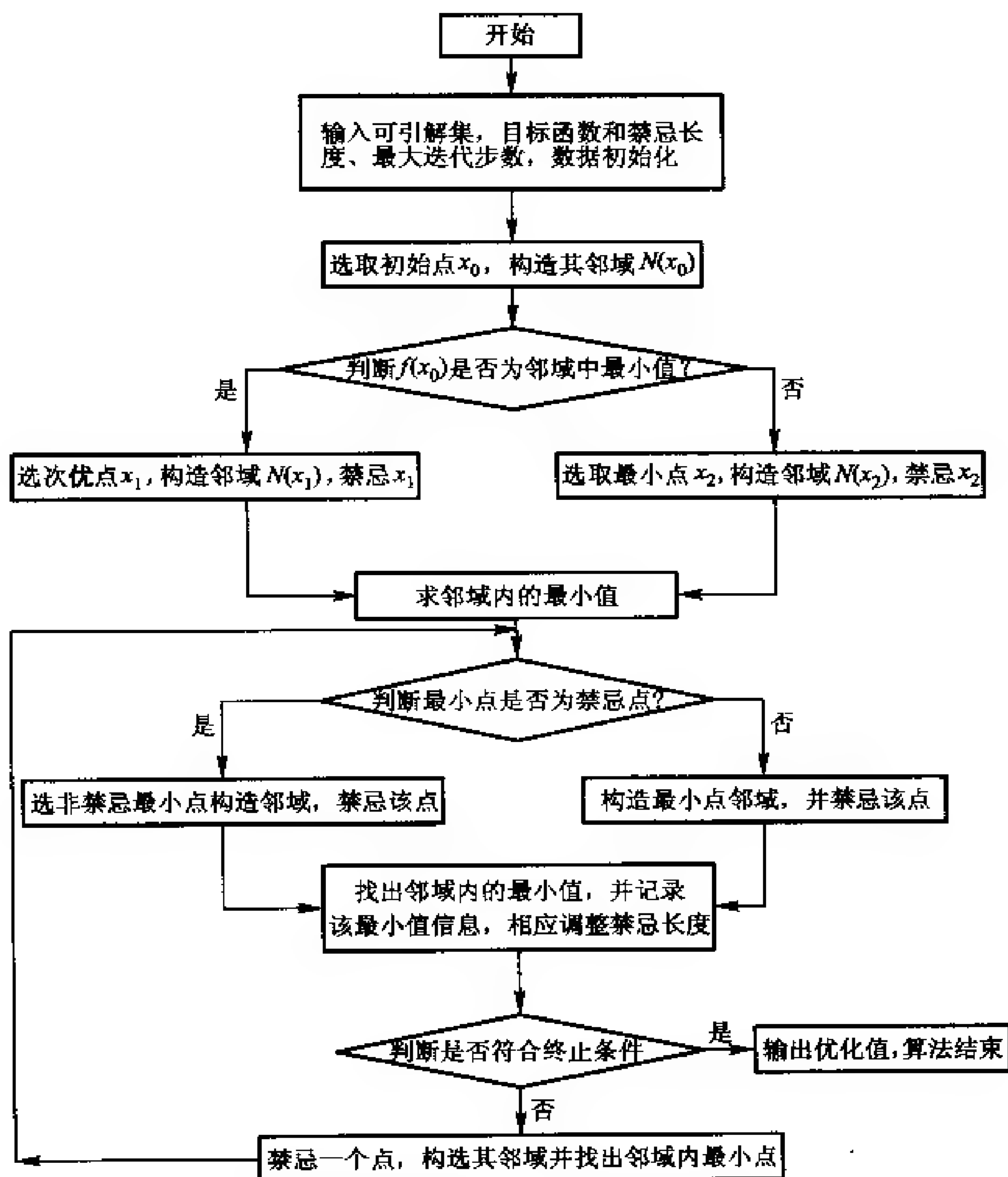


图 18－1

④ 输出计算结果，停止.

18.2.2　禁忌搜索算法的特征

禁忌搜索算法是一种人工智能算法，实现的技术问题是算法的关键.禁忌搜索算法的特征由禁忌对象和长度、候选集和评价函数、停止规则和一些计算信息组成.下面分别给予介绍.

1．禁忌对象、长度和候选集

禁忌对象和禁忌长度是禁忌表中的两个主要指标.禁忌对象是指禁

忌表中那些被禁忌的变化元素，由于解状态的变化分为解的简单变化，解向量分量的变化和目标值变化三种情况，所以禁忌对象的选取也相应的有对解的简单变化禁忌，对解向量的分量变化禁忌和对解的目标值变化禁忌三种情况.

(1) 解的简单变化禁忌

这种变化最简单，设 $x,y \in D$，其中 D 为优化问题的定义域，则当解从一个解 x 变化到另一个解 y 时，y 可能为局部最优解，为避开局部最优解，采用禁忌原则，当 y 的邻域 $N(y)$ 中有比它更优的解时，选择更优的解，当 y 为其邻域 $N(y)$ 的局部最优时，则选择比 y 更差的解.

(2) 解向量的分量的变化禁忌

当一个解向量 $x=(x_1,\cdots,x_{i-1},x_i,x_{i+1},\cdots,x_n)$ 采用与上面相同的禁忌原则，构造解 x 的邻域 $N(x)$ 各个分量在 x 的邻域内发生变化：

$$(x_1,\cdots,x_{i-1},x_i,x_{i+1},\cdots,x_n) \to (x_1,\cdots,x_{i-1},x'_i,x_{i+1},\cdots,x_n),$$

这是第 i 个分量发生变化的情况，向量的分量变化也包含有多个分量发生变化.

(3) 解的目标值的变化禁忌

在单目标值情况下，目标值变化禁忌类似于解的简单变化禁忌. 但多目标值情况下，需先通过综合评价转化为单目标，再按类似于解的简单变化禁忌处理.

针对三种不同状态的变化方式，禁忌的对象可以是任意一种. 实际应用时，要根据具体问题采用适当的一种方法. 解的简单变化比解向量的分量变化和目标值变化的受禁忌范围小，则相应增加了计算的时间，但也给予了较大的搜索范围. 解向量的分量变化和目标值变化的禁忌范围大，相应减少了计算的时间，但禁忌范围大容易陷在局部最优点. 因此，禁忌搜索算法的技术很强.

在算法构造和计算过程中，为了尽量少的占用机器内存，要求禁忌长度、候选集合尽量的小. 但禁忌长度过短造成了搜索的循环，候选集合过小造成过早地陷入局部最优.

禁忌长度是指被禁忌对象不允许被选取的迭代步数. 一般是给被禁忌对象 x 一个数 l（称为禁忌长度），要求 x 在 l 步迭代内被禁，在禁忌表中采用 $T(x)=l$ 记忆，每迭代一步，作运算 $T(x)=l-1$ 直到 $T(x)=0$ 时解禁. 有关禁忌长度 l 的选取，可归纳为如下的几种情况：

(1) l 为常数，如 $l=10$，$l=\sqrt{n}$，其中 n 为邻域中邻居的个数. 这种规则容易在算法中实现.

(2) $l \in [l_{\min},l_{\max}]$，此刻 l 是可变化的数，变化依据为被禁忌对象的

目标值和邻域的结构. 此时 l_{min}, l_{max} 为确定的，确定 l_{min}, l_{max}，一般根据问题的规模 N，限定变化区间 $[a\sqrt{N}, b\sqrt{N}]$, $0<a<b$，也可用邻域中邻居的个数 n 确定变化区间 $[a\sqrt{n}, b\sqrt{n}]$, $0<a<b$. 当给定了变化区间，则根据实际问题，实验和设计者的经验来确定 l 的大小.

(3) l_{min}, l_{max} 的动态选取. 有的情况下，用 l_{min}, l_{max} 的变化能达到更好的解.

总之，禁忌长度的选取与实际问题，实验和设计者的经验有非常紧密的联系，它决定了计算的复杂性. 禁忌长度选择过短会出现循环，过长又会使计算时间增长.

候选集合是由邻域中的邻居组成的，常规方法是从邻域中选择若干个目标值或评价值最佳的邻居入选. 若认为这样的计算量还是太大，则不在邻域的所有邻居中选择，而是从邻域中的一部分邻居中选择若干个目标值或评价值最佳的状态入选，也可以用随机选取的方法实现部分邻居的选取.

2. 评价函数

评价函数是候选集合元素选取的一个评价公式，候选集合的元素通过评价函数值来选取. 评价函数主要分为基于目标函数和其他方法两类：

(1) 基于目标函数的评价函数

通常用目标函数来作为评价函数，因为目标值是一个很直观的函数. 这一类主要包含以目标函数的运算所得到的评价方法，它的变形很多.

(2) 其他方法

但有时计算目标值比较复杂或耗时过多，为了方便和易于理解，常常会用其他函数来取代目标函数，作为替代的评价函数. 但替代的评价函数应该反映原目标函数的一些特性，如原目标函数对应的最优点还应该是替代函数的最优点，同时要注意减少计算的复杂性等.

3. 特赦规则

在禁忌搜索算法的迭代过程中，有时会出现候选集中的全部对象或者某一对象被禁忌，但如果解禁又会使其目标值有非常大的下降情况. 这时，为了达到全局的最优，可以让一些禁忌对象重新可选，这种方法称为特赦，相应的这种规则称之为特赦规则.

4. 记忆频率信息

在计算的过程中，记忆一些信息对解决问题是有利的. 如一个最好的目标值出现的频率很高，这时可以推测现有参数的算法可能无法再得到更好的解，因为重复的次数太高，有可能出现了多次循环. 这样根据解决问题的需要，可以记忆解集合，有序被禁对象组，目标值集合等的出现频率.

一般可以根据状态的变化将频率信息分为静态和动态两类.静态的频率信息主要指某些变化如解、对换或目标值在计算中出现的频率.动态的频率信息主要指从一个解、对换或目标值到另一个解、对换或目标值的变化趋势.

频率信息可以进一步加强禁忌搜索的效率.如可以根据频率信息动态控制禁忌的长度.当一个元素或一个序列重复出现,可以增加禁忌长度来避开循环.当一个序列的目标(评价)值变化较小时,增加该序列的每一个对象的禁忌长度,反之若变化较大时则减少.

5. 终止规则

禁忌搜索算法终究是一个启发式算法,所以不可能使禁忌长度充分的大,只希望在可以接受的时间里给出一个满意的解.这样需要有终止原则.大体上常用的终止原则有四种:

(1) 步数终止原则

这种终止原则是指给定一个充分大的数,使总的迭代次数不超过这个数.虽然这种原则便于操作,但无法保证解的效果.

(2) 频率控制原则

这种终止原则是指当某一个解、目标值或元素序列的频率超过一个给定的标准时,算法不改进,解已经不会改进,故终止计算.

(3) 目标值变化控制原则

这种终止原则是指如果在一个给定的步数内,目标值没有改变,则算法也不改进,解也已不会改进,故终止计算.

(4) 目标值偏离程度准则

这种终止原则是指对一些问题可以简单地计算出它们的下界 $\underline{z}$(目标为极小),若目标值为 $f(x)$,则对给定的充分小的正数 ε,当 $f(x)-\underline{z}\leqslant\varepsilon$ 时计算终止.

禁忌搜索算法虽然是人工智能在组合优化算法中的一个成功应用,但算法本身也存在着一些缺陷.首先,它对于初始解的依赖性较强,一个较好的初始解可以使算法在解空间中搜索到更好的解,但一个较差的初始解却会降低算法的收敛速度.其次,它在搜索过程中初始解只能有一个,迭代一次,也只能把一个解移动到另一个解.

18.3 模拟退火算法

模拟退火算法(简称 SA)是属于一种通用的随机搜索算法,是局部搜索算法的扩展.它的早期思想是在 1953 年由 Metropolis 提出来的,到

1983 年由 Kirkpatrick 等人成功地应用在组合最优化问题中.

模拟退火算法的基本思想来源于固体的退火过程.当加热固体到一定温度后,它的所有分子在状态空间自由运动.随着温度的逐渐下降,分子停留在不同的状态,分子运动逐渐趋于有序,最后以一定的结构排列.这种由高温向低温逐渐降温的过程称为退火.退火过程中系统的熵值不断减小,系统能量随温度的降低趋于最小值.固体退火过程与优化问题之间存在着类似性.Kirkpatrick 等人把 Metropolis 等人对用固体在恒定问题下达到热平衡过程的模拟引入到优化过程中.即如果

$$\Delta f = f[x(t+\Delta t)] - f[x(t)] \leqslant 0(\Delta t > 0),$$

则接受新状态,否则按概率 $P(\Delta f) = e^{-\frac{\Delta f}{T}}$ 接受新状态.$T = T(t)$ 为一随机时间 t 增加而下降的参变量,相当于退火过程中的温度.这种利用优化问题求解与物理系统退火过程的相似性,使用 Metropolis 算法,适当控制温度的下降过程,实现模拟退火,从而达到求解全局优化问题的随机性方法称之为"模拟退火算法".

模拟退火算法在搜索策略上引入了适当的随机因素和物理系统退火过程的自然机理,使得在迭代过程中出现可以接受使目标函数值变"好"的试探点,也可以以一定的概率接受使目标函数值变"差"的试探点.接受概率随着温度的下降逐渐减小.这样避免了搜索过程陷入局部最优解,有利于提高求得全局最优解的可靠性.因此,模拟退火算法具有实用范围广,求得全局最优解的可靠性高,算法简单,便于实现等优点.

18.3.1　模拟退火算法的基本原理

模拟退火算法利用模拟热力学中经典粒子系统的降温过程来求解规划问题的极值.Metropolis 过程的基本步骤为:

(1) 给定初始温度 T_0 及初始点 x_0,计算该点的函数值 $f(x_0)$;

(2) 随机产生扰动 Δx,得到新点 $x = x_0 + \Delta x$,计算新点的函数值 $f(x)$ 及函数值差 $\Delta f = f(x) - f(x_0)$;

(3) 若 $\Delta f \leqslant 0$,则接受新点,作为下一次模拟退火的初始点;

(4) 若 $\Delta f > 0$,则计算新点的接受概率 $P(\Delta f) = e^{-\frac{\Delta f}{KT}}$,产生[0,1]区间上的均匀分布的伪随机数 r,$r \in [0,1]$,若 $p(\Delta f) \geqslant r$,则接受新点为下一次模拟的初始点,否则放弃新点,仍取原来的点作为下一次模拟的初始点.

模拟退火优化算法就是按照一定的退火方案逐渐降低温度,重复

Metropolis 过程而构成的. 当系统温度足够低时,就认为达到了全局最优状态. 在模拟退火优化算法过程中,降温的方式对算法的影响很大. 如果温度下降过快可能就会丢失极值点,如果下降过慢,收敛速度又会大大降低. 为提高模拟退火优化算法的有效性分别提出了多种退火方案. 最典型的有经典退火方式和快速退火方式两种. 前者的降温公式采用 $T(t)=\frac{T_0}{\ln(1+t)}$,温度下降很缓慢,算法的收敛性也很慢;后者的降温公式采用 $T(t)=\frac{T_0}{1+\alpha t}$(其中 α 可改善退火曲线的形态),在高温区温度下降较快,而在低温区温度下降较慢,这符合粒子在高温时具有较低能量的概率比在低温小很多的热力学的分子运动理论,这样就可以把寻优的重点放在低温区.

18.3.2 模拟退火算法的基本步骤和实现的技术问题

模拟退火算法用一组冷却进度表参数来控制算法的进程,使算法在控制参数 t(随算法进程递减)徐徐"降温"并趋于零时,最终求得组合优化问题的相对全局最优解. 冷却进度表参数包括控制参数 t 的初始值 t_0、衰减函数、终止准则及 Markov 链长度 L_k,优化问题的一个解 i 及其目标函数 $f(i)$ 分别与固体的一个微观状态 i 及其能量 E_i 相对应. 令控制参数 t 担当固体退火过程中温度 T 的角色,对 t 的每一个取值,采用 Metropolis 接受准则,持续进行"产生新解—判断—接受或舍弃"的迭代过程,最终达到该温度下的"平衡点".

模拟退火算法如下所示:

算法 18.5

① 给定冷却进度表参数及其迭代初始解 x_0 和 $f(x_0)$;

② 参数 $t=t_k$ 时作 L_k 次试探搜索;

据当前解 x_k 产生一随机向量 z_k(对连续变量)或随机偏变量 q(对离散变最)得到 x_k 邻域的新的试探点 x'_k:

$$x'_k=\begin{cases}x_k+z_k, & \text{连续变量},\\ X(q+q'), & \text{离散变量},\end{cases}$$

其中 X 为离散变量的取值序列,q 为当前解的离散位置.

③ 产生一个在(0,1)上均匀分布的随机数 η,计算出在给定当前迭代点 x_k 和温度 t_k 下与 Metropolis 接受准则相对应的转移概率 p,有:

$$p=\begin{cases}1, & f(x'_k)\leqslant f(x_k),\\ e^{\frac{f(x_k)-f(x'_k)}{t_k}}, & f(x'_k)>f(x_k).\end{cases} \tag{18-1}$$

若 $\eta\leqslant p$，则接受新解，$x_k=x'_k$，$f(x_k)=f(x'_k)$，否则 x_k 不变.

④ 试探点搜索小于 L_k 次，转②，否则转③；

⑤ 满足迭代终止条件？

是，则停止，当前解为近似全局最优解，

否，则转⑥.

⑥ 根据给定的温度衰减函数，产生新的温度控制参数 t_{k+1} 及 Markov 链长度 L_{k+1}；

⑦ 重复②—⑥，直到找到最优解.

模拟退火算法是由式(18－1)描述的 Metropolis 准则既接受优化解也接受劣化解. 算法开始时 t 值较大，就可能接受较差的劣化解，但随 t 的减小，只会接受较好的劣化解. 最后在 t 值趋于零时，就不再接受劣化解了. 因此，模拟退火算法能从局部最优的陷阱中跳出来，以较大的概率获得全局最优解.

下面讨论算法实现中的一些技术问题，包括初始温度的选取、温度下降方法和停止温度的确定，每一温度 Markov 链的迭代步长和终止准则等.

1. 初始温度的选取

影响模拟退火算法全局搜索性能的重要因素之一就是初始温度的选取. 初始温度越高，算法搜索到全部最优解的可能性就越大，但同时所花费的计算时间也就越长. 初始温度低，虽可节省计算时间，但又会影响全局搜索性能，使搜索到的全局最优解的可能性减小. 通常在实际应用过程中，都要依据实验的结果对初始温度进行若干次的调整.

理论上，起始温度 t_0 应该保证平稳分布中的每一状态的概率相等，也就是要满足：

$$e^{-\frac{f_0}{t_0}}\approx 1,$$

由此得到初始温度的一个估计值　$t_0=k\delta$，

其中 k 为充分大的数，实际计算可根据问题特点确定，如选 $k=10,20,\cdots$ 实验值. 但注意 k 的选取过大会造成计算时间增加，k 过小又会使算法过早陷入局部最优点. δ 可按照 $\delta=\max\{f(j)\mid j\in D\}-\min\{f(j)\mid j\in D\}$ 选取，但 δ 有时会出现很难估计或很难精确知道最大值和最小值.

对于初始温度的选取可以通过取值计算估计和统计推断估计的方

法，通过数值计算可以估计出温度 t_0.

算法如下：

算法 18.6

① 给定一个常量 T；初始温度 t_0；γ_0（γ_0 接近于 1）；$R_0=0$. $k:=1$；

② 在这个温度迭代 L 步（L 为一个给定的常数）分别记录模拟退火算法中接受和被拒绝的状态的个数，计算接受的状态数同迭代步数 L 的比率 R_k；

③ 检查 $|R_k-\gamma_0|<\varepsilon$？

是，则停止计算；

否，当 $R_{k-1}, R_k<\gamma_0$，则 $k:=k+1, t_0:=t_0+T$，转②；

当 $R_{k-1}\geqslant\gamma_0$，且 $R_k\leqslant\gamma_0$，则 $k:=k+1, t_0:=t_0+\frac{T}{2}, T:=\frac{T}{2}$，转②；

当 $R_{k-1}\leqslant\gamma_0, R_k\geqslant\gamma_0$，则 $k:=k+1, t_0:=t_0-\frac{T}{2}$，转②；

对于 δ，通常采用统计的方法估计费用函数的上下限. 具体方法可参考文献[13].

2. 温度下降方法的确定

温度下降方法的确定是模拟退火算法中的另一个影响模拟退火算法全局搜索性能的重要因素.

在邻域搜索过程中，当解的质量变差的概率呈 Boltzmann 分布时，对数降温方式可以使模拟退火算法收敛于全局最优解，即：

$$t(k)=\frac{K}{\ln(1+k)},$$

其中 k 为正的常数，k 为降温次数.

但是当解的质量变差的概率呈 Cauchy 分布时，对数降温方式可以使模拟退火算法收敛于全局最优解，即

$$t(k)=\frac{K}{1+k}.$$

在实际问题的应用中，通常采用两种非常直观的下降方法，一种是每一步温度以相同的比率下降，即 $t_{k+1}=\alpha t_k$，其中 $k\geqslant 0, 0<\alpha<1$，α 为降温系数，α 越接近 1 温度下降越慢. 由于这种方法简单易行，应用很广；另一种是每一步温度以相同的长度下降，即：$t_k=\frac{t_0(N-k)}{N}$，其中 t_0 为初始温度，N 为温度下降的总次数. 这一下降的方法的优点是易于操作，且可以简单地控制温度下降的总步数.

3. 每一温度迭代长度的确定

每一温度迭代长度也是影响模拟退火算法全局搜索性能的重要因素，在实际计算中要根据问题的特点确定合理的迭代长度，常用的方法有如下几种.

(1) 固定迭代步数

这种方法很简单，就是在每一温度确定相同的迭代步数，步数的选取同实际问题的大小有关，一般采用与邻域大小直接相关的规则.

(2) 由接受和拒绝的比率来控制迭代步数

在温度高时，每一状态被接受的概率基本相同，几乎所有状态都被接受，这时可以使同一温度的迭代步数尽量的少，当温度逐渐降低以后，越来越多的状态被拒绝，为避免为早的陷入局部最优，可以相应地增加迭代的步数. 一般常用的方法是给定一个充分的步长上限 U 和一个接受次数上限 W，当某一温度的实际接受次数等于 W 时，在此温度下不再迭代而使温度下降，否则继续迭代到上限步数 U. 另一种方法是再给定一个接受比率 r 和一个迭代步长下限 V，每一温度至少迭代 V 步，当迭代步数超过 V 步时，如果同一温度迭代被接受次数同总次数的比率不小于 r，则不再迭代而使温度下降，否则继续迭代到上限步数 U. 当然也可以用拒绝次数为指标得到一些类似的控制迭代步数的规则.

另外，还有概率控制法等其他方法.

4. 算法的终止准则

模拟退火算法的终止规则主要采用一些比较直观的方法，如：

(1) 零度法

模拟退火算法的最终温度为零，所以常给定一个比较小的正数 ε，当温度 $t_k \leqslant \varepsilon$ 时，算法终止，表示达到了最低温度.

(2) 循环总数控制法

这种方法是设总的温度下降次数为一定值 N，当温度迭代次数达到 N 时，算法终止. 这一原则可分为两类，一类是整个算法的总迭代次数为一定值，另一类是温度下降次数为一定值.

(3) 基于不改进规则的控制法

这种方法是指在一个温度和给定的迭代次数内没有改进当前的局部最优解，则算法终止.

(4) 接受概率控制法

模拟退火算法的基本思想是跳出局部最优解，如果在较高温度时没能跳出局部最优解，在较低温度时能跳出局部最优解的可能性也更小，所以就给定一个比较小的数 p，在一个温度和给定的迭代步数内，除当前局

部最优解外,其他状态接受概率都小于 p 时,算法终止.

另外,还有一些算法终止准则,如邻域法,Lundy 和 Mess 方法,Aarts 和 VanLaarhoven 法等等.

分析模拟退火算法的基本原理看出,理论上是用一个 Markov 链描述模拟退火算法的变化过程,所以具有全局最优性,在具体应用模拟退火算法中发现,它也有许多不足,主要是有许多参数需要人为调整,如起始温度、温度下降方案、固定温度时的迭代长度、终止规则等等,这样就造成了计算结果的差异,因而很难保证计算结果为全局最优点.这就需要通过大量的数值模拟计算来选择好的参数.

18.4　遗传算法

遗传算法(简称 GA)是在 1975 年首次由 Holland 提出来的.它的基本思想来源于 Darwin 的进化论和 Mendel 的遗传学说.算法主要借用生物进化论中"适者生存"的规律,即最适合自然环境的群体往往产生更大的后代群体.它借鉴了生物进化的一些特性,是一种新的全局最优搜索算法,简单实用,适合于并行处理,在优化调度、运输问题、组合优化等邻域已得到广泛应用,成为求解任意函数优化问题的强有力的工具.

遗传算法可以解决组合优化问题以及目标函数或某些约束条件下不可微的非线性优化问题.它以适应度函数为依据,通过对群体中的个体施加遗传操作,实现群体内个体结构的重组,使群体内个体一代代得以优化并逐渐接近全局最优解.

18.4.1　遗传算法的基本原理和步骤

对于一般的优化问题

$$\begin{aligned} \min \quad & w = f(x,y,z), \qquad w \in \mathbf{R}, \\ \text{s.t.} \quad & G(x,y,z) \geqslant 0, \quad (x,y,z) \in \Omega. \end{aligned}$$

遗传算法包含以下主要的处理步骤,首先将优化问题的一组基本可行解 (x,y,z) 进行编码,表示一组二进制的字符串.一个解的编码称为一个染色体,组成编码的元素称为基因,编码主要是用于优化问题解的表现形式以及后面遗传算法中的计算,开始一般总是随机产生一些个体(即初始解),根据确定的优化问题的目标函数(称为适应函数),对每个个体进行评价,得出个体的适应函数值,由适应函数值的大小决定的概率分布来确定哪些染色体适应生存,哪些被淘汰.然后选择出来的个体经过交换和变异进行再组合,产生一个新个体.新个体产生过程中可能发生基因变

异，变异使某些解的编码发生变化，使解有更大的遍历性. 新个体继承了上一代的优良特性. 遗传算法就是通过对生物基因的复制、交换和变异这三种模拟来实现其优化过程的.

复制是遗传算法的关键，它来源于自然界中适者生存的思想，是指按一定的概率从群体中选择若干的染色体，作为双亲用于繁殖后代，产生新的染色体加入下一群体中. 生物进化的适者生存规律使得最具有生存能力的染色体以最大的可能生存.

交换体现了自然界中信息反馈的思想，是指对于选中的用于繁殖的每一对染色体，随机地选择同一整数，将双亲的基因码链在此位置相互交换.

变异体现了自然界中生物会偶然发生基因变异的思想，是指以一定的概率从群体中随机的选取若干染色体，对于选中的染色体，随机选取某一位进行取反运算.

对某一代种群经过对生物基因的复制、变换和变异，将产生新一代种群，再重复此过程，直到群体或最优点的性能达到满意程度. 遗传算法的基本步骤为：

算法 18.7

① 选择问题的一个编码，给出一个有 N 个染色体的初始群体 $p(1)$，当前进化步数 $t := 1$；

② 计算 $p(t)$ 中的每一个个体 $p_i(t)$ 的适应函数 $f_i = \text{fitness}(p_i(t))$；

③ 若终止规则满足，则算法停止，

否则，计算概率：

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j}, i = 1,2,\cdots,N, \tag{18-2}$$

并以概率分布(18－2)从 $p(t)$ 中随机选一些染色体构成一个种群.

$$\text{new } p(t+1) = \{p_j(t) \mid j = 1,2,\cdots,N\};$$

④ 通过交配，交换概率为 p_c，得到一个有 N 个染色体的 $\text{crossp}(t+1)$；

⑤ 以一个较小的概率 p，使得一个染色体的一个基因发生变异，形成 $\text{mutp}(t+1)$；$t := t+1$，一个新的群体 $p(t) = \text{mutp}(t)$，转②；

18.4.2 遗传算法的技术问题

遗传算法的主要问题是算法如何实现的技术问题. 归结起来有如下一些因素：

1. 解的编码和解码

解的编码是遗传算法的最基础工作，只有在编码之后才可能有其他的计算. 算法的最后一个工作则是通过解码得到问题的一个解.

比较直观和常规的方法是 0－1 二进制编码，称为常规码，它与人类的染色体成对结构类似. 这种编码算法可以使算法的算子构造比较简单. 对于一些优化问题这样表示很直观、简便. 但在表示有些组合优化问题时会显得无效或不方便. 所以除此之外，还有非 0－1 码，称为非常规编码，这与实际问题联系紧密.

2. 初始群体的选取

在计算开始时，需要产生一些待优化问题的可能解，称为初始群体. 初始群体可用随机方式产生，也可以用其他的一些启发式算法或经验选择，主要针对实际问题而定.

3. 群体规模的确定

群体规模在算法第一层就需要确定，实际应用中时常取个体编码长度数的一个线性倍数. 当然群体规模的选取也可以是变化的，当多个进化代没能改变解的性能，保持现有的群体规模已很难改进解时，可扩大群体的规模. 反之，若解的改进已经非常好时，就可以减少群体规模，使计算速度加快.

4. 适应函数的确定.

一般情况下，适应函数同目标函数相关，以此保证较优的解有较大的生存机会，适应函数有以下几类：

(1) 简单适应函数

简单适应函数是目标函数的简单变形，它的优点是构造简单，与目标函数直接相关. 但它的缺点是可能使算法在迭代过程中出现收敛到一些目标值近似的不同染色体而难以区别.

(2) 加速适应函数

加速适应函数有非线性加速适应函数，线性加速适应函数等. 它们的思想是希望开始时每一个状态有比较大的选取性，随着计算的步步进行，逐渐拉开目标值不同对应状态的档次.

(3) 排序适应函数

排序适应函数可以避开对目标函数进行线性、非线性等加速适应函数的早熟可能，使每一代当前最好的解以最大的概率遗传.

5. 三个算子

遗传算法的三个算子是种群选取、交配和变异. 前面遗传算法中新群体产生中的一个主要问题就是如何选取种群. 种群选取通常有这样一些策略：

（1）根据适应度比例或排名的方法

种群是以一个概率分布——轮盘赌的形式选择个体而产生的，如首先计算各个个体的相应适应度：$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j}$，$i = 1,2,\cdots,N$，以 p_i 为选择概率将一个圆盘分成 N 份，则第 i 个扇形的中心角为 $2\pi p_i$，若某个参照点落在第 i 个扇形内，则选择个体 i. 显然这样相对适应度高的个体被选中的机会比相对适应度低的个体机会更大，它的基因结构被遗传到下一代的可能性也就大. 也可以根据个体适应度在群体的排名来确定选取概率 p_i，再由上述方法选取.

（2）根据局部竞争机制的方法

这种方法是先随机地在群体中选择 k 个个体进行比较，然后选取适应度好的个体作为下一代的父体. 显然这种方法可以使适应度的个体有较大的遗传机会.

（3）最优保存选取方法

这种方法常结合其他选取方法使用. 它的具体方法是保存当前群体中的适应度最高的个体，而不让它参与交叉和变异运算，而用来替换本代群体中经过交配、变异后所产生的适应度最低的个体. 从而既防止遗传操作过程中所获得的阶段最优个体不被随机性破坏，同时也可能防止某个局部最优个体不易被淘汰从而降低算法的全局搜索能力.

总之，选取方法要根据问题的特点和群体规模灵活采用或相互配合使用.

交配是基因算法获得优良个体的重要手段，就是从代码群中随机地选取两个作为父代进行杂交，杂交过程中. 两个父代部分结构加以替换重组而生成新的个体，这样父代的部分码遗传给子代，使优良的特性一代代地遗传下去. 具体交配的方法有很多，以常规码来说，主要的交配方法有两大类，一类是双亲遗传法，如：双亲双子法、变化交配法、多交配位法、双亲单子法、显性遗传法等等，另一类是单亲遗传法. 下面简要说明一下：

（1）双亲双子法

首先在确定双亲以后，以一个随机位进行位之后的所有基因的对换，然后形成两个后代，即如父代 A = 100/100，父代 B = 010/010，则形成子代 A = 100/010 和子代 B = 010/100.

（2）变化交配法

但有些双亲，采用上述方法可能导致父代与子代完全相同，从而影响收敛速度和搜索范围. 这时就从头开始先比较它们相同的基因，从不同基

因位置按前面方法随机选交配位. 如父代 A = 1101001，父代 B = 1100010，若在前三位选为交配位，交配不会有变化，但是如果在第 4 位以后随机选取，交配就会有变化了.

(3) 多交配位法

随机选择多个交配位，双亲以一个交配位到下一个交配位基因相互替代和下一个交配位到再下一个交配位不变，交叉形成两个新的后代，如父代 A = 11/01/001 和父代 B = 11/00/010，则形成子代 A = 11/00/001 和子代 B = 11/01/010.

(4) 双亲单子法

这是使一对双亲只有一个后代，既可以从常规交配法的两个后代中随机选择一个，也可以根据优胜劣汰从两个后代中选一个好的. 还可以把双亲单子法与变化交配法和多交配位法结合使用.

(5) 显性遗传法

对双亲中的基因，有些是具有优超关系的，这些基因必将遗传到下一代，如父代 A = 1101001，父代 B = 0000010，则形成子代 1101011.

(6) 单亲遗传法

单亲遗传法与双亲遗传法的区别是它只有一个父代，下一代的产生通过单亲自身的基因变化. 如选定一个单亲，随机选两个基因位置，将两个基因位置的元素进行交换或选定两个交配位，通过交配位之内的基因倒排得到子代.

另外，前面说到，这些方法是对于常规码的方法，对于非常规码，这些交配方式都失效了. 如对 n 个整数排列的非常规码的交配方法，主要有非常规码的常规码交配法和不变位法两种方法.

变异能够导入新的遗传信息，使寻优有可能指向未探知区域，是提高全局最优搜索能力的有效步骤. 同时，也能保持群体差异，防止过早出现收敛现象. 常见的变异算子有逆转变异、对换变异和插入变异. 逆转变异是在个体编码中随机选择两点，再将这两点内的子串按反序插入原位置中. 对换变异是随机选择子串中的两点并交换其数码. 插入数码是从字串中随机选择一个码，将此码插入随机选择的插入点中间. 其他的还有连续逆转变异、进化变异和混合变异等.

遗传算法尽管能解决任意函数、高维空间的组合优化问题，但由于在进化搜索过程中，每代要维持一定规模的群体，而群体大小的选择会影响算法的使用，群体规模过大，包含的信息量也会大，使得计算量也大，不利于算法实现，群体规模过小，包含的信息量也小，又使得算法得不到充分发挥. 另外由于遗传算法中最重要的遗传算子——交叉算子使群体中的

染色体具有局部相似性,以及变异概率太小的原因,也使搜索转向其他解空间进行搜索的可能性很小.这些都使得遗传算法的应用受到限制.

18.5　神经网络算法

人工神经网络(简称 ANN)是由人工建造的,以有向图为拓扑结构的动态系统,通过对连续或离散的输入作状态响应而进行信息处理.实质上就是模仿人的大脑的工作方式而设计的,具有大量连接的并行分布式处理器,它具有通过学习获取知识并解决问题的能力,并且知识分布存储在连接权中.它在一些工程实际和科学研究中,已经显示了很大的威力,尤其是在控制优化、金融预测和管理、信号处理、模式识别与图像处理等领域得到非常广泛的应用.

自 20 世纪 40 年代美国生理学家 W. S. Mcculloch 和数学家 W. A. Pitts 建立第一个数学模型——二值神经元模型(M - P 模型)以来,ANN 的发展几经兴衰,先后提出了几个有代表性的模型:感知器模型(F. Rosenblatt,1957 年)、自适应线性神经元 Adaline 模型(Widrow,1961 年)等.

ANN 有单层和多层之分,每一层包含若干神经元,即信息处理元,各神经元之间用带可变权重的有向弧连接,网络通过对已有信息的反复学习训练,逐渐调整改变神经元连接的权重,最终达到处理信息、模拟输入输出关系的目的.神经网络技术与传统的数据处理方法相比,在处理模糊数据、随机性数据、非线性数据等方面具有明显的优越性,特别对大规模、结构复杂、信息不明确的系统非常适用.因为它只需知道能引起输出变化的非恒定性因素.而不需要很多其他参数,如输入输出之间的确切关系等参数.

18.5.1　人工神经网络的基本概念

人工神经网络模型是基于生物学中的神经网络的基本原理而建立的.大脑的一个重要成分是神经网络.神经网络由相互关联的神经元组成.每一个神经元由内核、轴突和晶枝组成.晶枝形成一个非常精密的"毛刷"环绕在内核周围.轴突就像一根管道,通过它将内核的信息传递给其他内核的晶枝.一个神经元通过晶枝接收到一定信息后,对信息进行处理,再通过它所控制的突触传递给其他的神经元.

神经元分"抑制性"或"兴奋性"两种.当一个神经元的晶枝接收的兴奋性信息累计超出某一值时,这个固定值称为阈值.如神经元接收到其他

神经元传递的信息被激活，并传递出一个信息给其他神经元，这种神经元为“兴奋性”的；如神经元虽然接收到其他神经元传递的信息，但没有向外传递信息，称神经元为“抑制性”的.

首先来看二值神经元模型（M－P 模型）.

假设一个神经元通过晶枝接收到 n 个信息. 如图 18－2 所示.

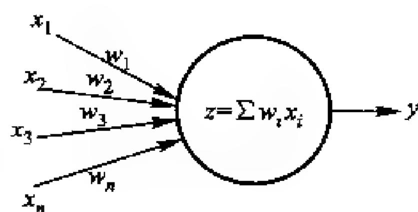


图 18－2

w_i 表示神经元对第 i 个晶枝接收到信息的感知能力，称为关联权. 函数 $f(z)$ 称为输出函数（激活函数）. M－P 模型的输出函数定义为

$$y = f(z) = \operatorname{sgn}\left(\sum_{i=1}^{n} w_i x_i - \theta\right), \qquad (18-3)$$

其中 $\operatorname{sgn}(x) = \begin{cases}1, & x > 0, \\ 0, & 其他,\end{cases}$ θ 称为阈值.

由式（18－3）可知，当 w_i 固定时，对给定的一组输入 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$，可计算得到输出值. 要使计算输出的值与实际值吻合，关键是确定参数 w_i. 人工神经网络的主要工作就是建立模型，并确定 w_i 的值. 根据人工神经网络模型建立和权值 w_i 的确定方法的不同，神经网络模型分为前向型（也称前馈型）人工神经网络和后向型（也称后馈型）人工神经网络.

前向型人工神经网络的特点是将神经元分为层，每一层的神经元之间没有信息交流，并且一层一层地同步计算，也就是说，各神经元接受前一层的输入，并输出给下一层，没有反馈. 结点分为两类，即输入单元和计算单元，每一计算单元可有任意个输入，但只有一个输出.

后向型人工神经网络的特点是将整个网络看成是一个整体，神经元之间的信息传递关系不再是从一层到另一层，而是各神经元之间存在着联系. 也就是说，所有结点都是计算单元，同时也可接受输入，并向外界输出.

下面各节分别讨论.

18.5.2 人工神经网络的基本模型

1. M－P 模型

前面提到，最早提出的神经元模型是 M－P 模型，M－P 模型是大多数神经网络模型的基础. 但 M－P 模型在发表时没有给出一个学习算法来调整神经元之间的连接权，但可以是用一些常见算法，如 Hebb 学习规则来调整神经元连接权，以达到学习的目的. Hebb 学习规则调整 w_{ij} 的原

则为：若第 i 和第 j 个神经元同时处于兴奋状态，则两者之间的连接应当加强，即

$$\Delta\omega_{ij} = \alpha u_i u_j,$$

其中 α 表示学习速率的比例常数.

2. 感知器模型

若只有一个输入层与一个输出层称为简单感知器.简单感知器模型仍然是 M－P 模型的结构，两者的区别在于神经元之间的连接权的变化.感知器的连接权定义为可变的，这样就被赋予了学习的功能.

感知器的处理单元对 n 个输入进行加权和操作，即

$$y = f\left(\sum_{i=1}^{n} w_i x_i - \theta\right), \tag{18-4}$$

其中 w_i 指第 i 个输入到处理单元的连接权值，θ 为阈值.感知器的输出 f 通常定义为阶跃函数：

$$y = \begin{cases} 1, & \sum_{i=1}^{n} w_i x_i - \theta > 0, \\ 0, & \sum_{i=1}^{n} w_i x_i - \theta \leqslant 0. \end{cases}$$

感知器的学习是通过监督学习过程来实理的.样本矢量 X 输入到感知器中，感知器中的各处理单元按（18－4）式来计算，得到实际输出 y，将其与期望输出 d 的差来修正下一计算步的权值.简单感知器的学习算法称为误差学习算法.

算法 18.8

输入给定值 k，

① 选择一组初始权值 $w_i(0)$；

② 计算误差 $\delta = d - y$；

③ 若 $\delta < k$？

是，停止计算，

否，转④；

④ 更新权值（阈值可视为输入恒为1的一个权值）

$$\Delta\omega_i(t+1) = \omega_i(t+1) - \omega_i(t) = \eta(d - y(t))x_i;$$

⑤ 重复②—④，直到满足要求.

式中 η 为学习步长，它是区间(0,1)上的一个常数，取值与训练速度和 w 收敛的稳定性有关，d,y 为神经元的期望输出和实际输出，x_i 为神经元的第 i 个输入.

如果在输入和输出层间加上一层隐含层的神经元，称为多层感知器，则构成多层前向网络. 多层感知器只允许调节一层的连接权，这样无法给出一个有效的多层感知器学习算法.

3. 自适应线性神经元

自适应线性神经元是一种连续时间线性网络，主要用于自适应系统等一些连续可调过程.

模型通过输入理想响应 d_k，用类似于误差学习的算法 LMS 算法来训练神经元. 该方法通过调整单个神经元的权值，使神经元的实际输出与理想输出的差为最小.

LMS 算法是采用误差最速下降准则，使用的误差测度函数为均方误差. 系统权向量可从任意初始值开始，然后计算均方误差函数的梯度，权值依负梯度变化. 不断重复此过程，使平均误差不断减少，达到优化目的. 即

$$w_{k+1} = w_k + \mu(-\nabla_k), \tag{18-5}$$

其中 w_k 为当前权值，w_{k+1} 为下一步权值，μ 为学习参数，∇_k 为均方误差曲面上对应于 $w = w_k$ 的梯度值.

设均方误差梯度函数为：

$$E = \sum_k \left(\frac{1}{2}\left(\sum_j \left(d_j^k - \sum_i w_{ij}x_i^k\right)^2\right)\right), \tag{18-6}$$

其中，d_j^k 为在 j 输出单元的期望输出，第 j 单元的实际输出为 $y_i = \sum_i w_{ij}x_i^k$.

(18－6)可记为　$E = \sum_k \left[\frac{1}{2}(d_k - w_k x_k)^2\right]$，

其中 x_k 为当前模式输入向量.

定义瞬时速度为　$\nabla_k \overset{\Delta}{=} \frac{\partial E_k}{\partial w_k} = -(d_k - w_k x_k)x_k = -\varepsilon_k x_k$，

由(18－6)知　$w_{k+1} = w_k + \mu(d_k - w_k x_k)x_k$，

即

$$\Delta w_k = \mu\varepsilon_k x_k. \tag{18-7}$$

由于是采用瞬时梯度下降来训练网络权值，所以 LMS 算法实际上是一种近似的梯度下降，它偏离了 E 真正的梯度下降方向，但这种偏离当学习速度 μ 很小时可以忽略不计.

若要实现全局误差函数 E 上的梯度下降，则 w_k 的变化应正比于负平均梯度. 即

$$\Delta w_k = -\mu\frac{\partial E}{\partial w_k} = \sum_{k=1}^{m}\left(-\mu\frac{\partial E_k}{\partial w_k}\right), \tag{18-8}$$

LMS 算法的步骤为：

$$w_{k+1} = w_k + \frac{\alpha}{|x_k|^2}\varepsilon_k x_k, \tag{18-9}$$

其中，ε_k 为当前误差，α 为学习步长，一般取为 $0.1 < \alpha < 1.0$，$|x_k|^2$ 为权值个数.

记
$$\varepsilon_k = d_k - x_k^T w_k,$$

则
$$\Delta\varepsilon_k = \Delta(d_k - x_k^T w_k) = -x_k^T \Delta w_k,$$

由(18－9)有
$$\Delta w_k = w_{k+1} - w_k = \frac{\alpha}{|x_k|^2}\varepsilon_k x_k,$$

故
$$\Delta\varepsilon_k = -x_k^T \frac{\alpha}{|x_k|^2}\varepsilon_k x_k = -x_k^T x_k \frac{\alpha}{|x_k|^2}\varepsilon_k = -\alpha\varepsilon_k.$$

下面给出 LMS 学习算法步骤：

算法 18.9

① 初始化 w,α,d_k,δ；

② 选择一输入样本 x；

③ 计算输出 y；

④ 计算 $\varepsilon < \delta$?

是，停止计算，

否，转⑤；

⑤ 计算 Δw.

18.5.3　前向型人工神经网络

前向型人工神经网络的计算一般分为两个阶段：学习阶段和工作阶段. 第一阶段是学习阶段，此时各计算并单元状态不变，各连线上的权值可通过学习来修改，所以这个阶段的主要工作就是：在选择网络模型和学习规则后，根据已知的输入和输出学习数据，通过学习规则确定神经网络的权值 $w_1,w_2,\cdots,w_n$，这个阶段的输入是学习数据中的输入数据，通过人工神经网络的输出同学习数据中的理想输出数据的比较，确定神经网络中的权值；第二个阶段是工作阶段，此时单元状态变化，在权值确定的基础上，用带有确定权值的神经网络去解决实际问题. 所以这个阶段的主要工作是：根据第一个阶段确定的模型和得到的权值 ，在输入实际问题的输入数据后，得出一个结论.

依据前向型神经网络的层数又可分为单层前向神经网络和多层前向人工神经网络.

1. 单层前向神经网络

单层前向神经网络模型如图 18－3 所示：

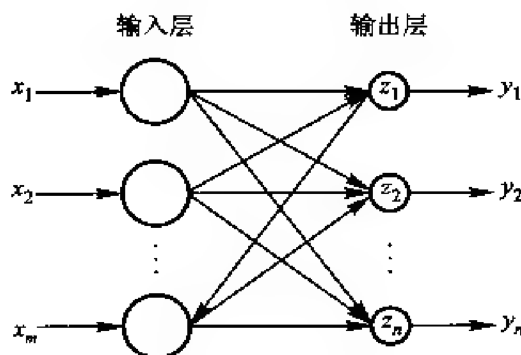


图 18－3

在单层前向神经网络模型中，只有一个输入层和一个输出层，输入层中的每个元素表示轴突，输出层中的每一个元素代表一个神经元的内核.单层前向神经网络中各层中的元素无权值相连，输出层通过输入层的输入数据与理想输出结果的比较来确定权值，再用确定了权值的神经网络应用到实际问题中.

记输入变量为 $\boldsymbol{x}=(x_1,x_2,\cdots,x_m)^{\mathrm{T}}$，输出结果为 $\boldsymbol{y}=(y_1,y_2,\cdots,y_m)^{\mathrm{T}}$，阈值 $\boldsymbol{\theta}=(\theta_1,\theta_2,\cdots,\theta_m)^{\mathrm{T}}$，$w_{ij}$表示输入 i 与神经元 j 的权值，则第 j 个神经元接收到的值为：

$$y_j = f\left(\sum_{i=1}^{m} w_{ij}x_i - \theta_j\right),\ j = 1,2,\cdots,n, \tag{18－10}$$

记

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix}$$

矩阵表示为

$$\boldsymbol{y} = f(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{\theta}).$$

依据激活函数的类型，可以把单层前向型神经网络分为线性网络和非线性网络.

如果一个神经网络的激活函数是线性函数，满足：

$$f(cz) = cf(z), \tag{18－11}$$

$$f(x+y) = f(x) + f(y),$$

其中 c 是一个实常数，x,y,z 为实变量值，则称之为简单神经网络.

如简单神经网络如果满足 $\quad \boldsymbol{y} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} \quad$ (18－12)

时，每一个神经元满足

$$y_i = (w_{1j}w_{2j}\cdots w_{mj})(x_1,x_2,\cdots,x_m)^{\mathrm{T}}. \tag{18－13}$$

从(18－13)式可知，每个神经元的输出只与其相关的权值和输入有

关，因此，n 个神经元是相互独立的，所以可以采用单个神经元研究其神经网络的特性.

如果一个神经网络的激活函数是非线性函数时，则称之为非线性神经网络. 常见的非线性函数有以下几种：

（1）符号函数：$f(z)=\begin{cases}1 & (z>T)\\ 0 & (其他)\end{cases}$；

（2）Sigmold 形函数：$f(z)=\dfrac{1}{1+e^{-z}}$.

神经网络具有学习功能，这种学习归结为神经元连接权的变化. 我们需要采用一些常见的算法来调整神经元连接权，以达到学习的目的. 常用的学习算法如 Hebb 规则和最小二乘学习规则等.

2. 多层前向神经网络

目前人工神经网络应用较多的是 BP 网络模型，BP 神经网络是一种多层前馈神经网络. 这是 1985 年 Rmelhart 等提出的，它系统地解决了多层神经元网络中隐单元层连接权的学习问题.

多层神经网络模型的拓扑结构如图 18－4 所示. 它由输入层、中间层（隐层）和输出层组成. 中间层可以是一层或多层的. 相邻层各种神经元间形成完全连接关系，而同一层内各种神经元之间没有任何连接关系. m 个输入信号从输入层进入网络，经激活函数变换后到达隐层，然后再经激活函数变换到达输出层构成 n 个输出信号，所以 BP 神经网络是一个从 m 维空间 $\mathbf{R}^m$ 的输入到 n 维向量 $\mathbf{R}^n$ 的输出的高度非线性映射.

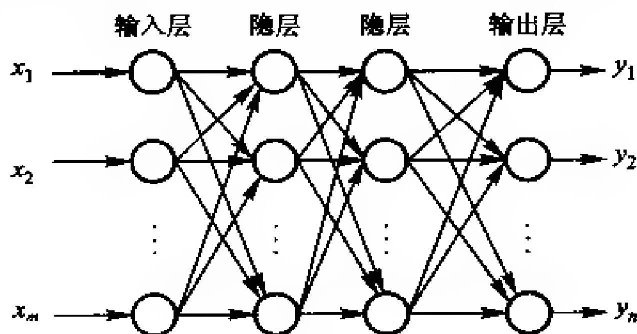


图 18－4

BP 网络的学习过程分为正向传播和反向传播两部分. 当正向传播时，输入信号从输入层经隐含层处理后传向输出层，每一层神经元的状态只影响下一层的神经元状态. 若输出层得不到希望的输出，则转入反向传播，将误差信号沿原来的神经元连接通路返回，返回过程中，逐一修改各层神经元连接的权值.

对于图 18－4 的前向多层网络，设有 p 个训练样本，即 p 个输入－输出对 $(\boldsymbol{x}_k,\boldsymbol{T}_k)$，$(k=1,2,\cdots,p)$. 其中，$\boldsymbol{x}_k$ 为第 k 个样本输入向量 $\boldsymbol{x}_k=(x_{k1},x_{k2},\cdots,x_{kM})^{\mathrm{T}}$，$M$ 为输入向量维数；$\boldsymbol{T}_k$ 为 k 个样本期望输出向量：$\boldsymbol{t}_k=(t_{k1},t_{k2},\cdots,t_{kN})^{\mathrm{T}}$，$N$ 为输出向量维数. $\boldsymbol{o}_k=(o_{k1},o_{k2},\cdots,o_{kN})^{\mathrm{T}}$ 为网络的实际输出向量. w_{ji} 为前一层第 i 个神经单元输入到后一层第 j 个神经单元的权重.

当神经元为输入层单元时，$\boldsymbol{o}_k=\boldsymbol{x}_k$.

现采用 S 型函数作为激发函数，对第 k 个样本第 j 个网络神经元 u_j 的状态定义为：

$$Net_{kj}=\sum_i w_{ji}o_{ki}+\theta_j,$$

则第 j 个神经元的输出为：

$$o_{kj}=f(Net_{kj})=\frac{1}{1+\mathrm{e}^{Net_{kj}}}=\frac{1}{1+\exp\left(-\sum_i w_{ji}o_{ki}-\theta_j\right)},$$

其中，θ_j 为单元 u_j 的阈值. 此时

$$f_1'(Net_{kj})=\frac{\partial o_{kj}}{\partial Net_{kj}}=o_{kj}(1-o_{kj}),$$

则对输出层单元 $\qquad \delta_{kj}=(t_{kj}-o_{kj})o_{kj}(1-o_{kj}),$

对隐层单元 $\qquad \delta_{kj}=o_{kj}(1-o_{kj})\sum_m \delta_{km}w_{mj},$

权值调节为 $\qquad \Delta w_{ji}(t+1)=\eta\delta_{kj}o_{ki},$

其中 η 为学习速率，η 是按梯度搜索的步长，η 越大，权值的变化越剧烈.

BP 算法把一组样本的输入输出问题变为一个非线性优化问题，使用了优化中最普通的梯度下降法. 采用了迭代运算求解权值. 通过加入隐层节点单元使优化问题的可调参数增加，从而得到更精确的解. 算法的实现步骤为：

算法 18.10

① 置各种权值或阈值的初始值：$w_{ji}(0)$，$\theta_j(0)$ 为小的随机数值；

② 提供训练样本，输入矢量 $\boldsymbol{x}_k$，期望输出 d_k；

③ 计算网络的实际输出及隐层单元的状态：

$$o_{kj}=f_j\left(\sum_i w_{ji}o_{ki}+\theta_j\right);$$

④ 计算训练误差

$$\delta_{kj}=\sigma_{kj}(1-o_{kj})(t_{kj}-o_{kj}) \qquad \text{（输出层）},$$

$$\delta_{kj}=\sigma_{kj}(1-o_{kj})\sum_m \delta_{km}w_{mj} \qquad \text{（隐含层）};$$

⑤ 修正权值和阈值

$$w_{ji}(t+1) = w_{ji}(t) + \eta\delta_j o_{kj} + \alpha[w_{ji}(t) - w_{ji}(t-1)],$$
$$\theta_j(t+1) = \theta_j(t) + \eta\delta_j + \alpha[\theta_j(t) - \theta_j(t-1)];$$

⑥ 对下一输入样本重复②—⑤，判断指标是否满足精度要求：$E \leqslant \varepsilon$，ε 为精度.

BP 模型虽然应用很广泛，但 BP 算法本身也有不足. 如它学习算法的速度慢，网络隐层节点数的选取有盲目性；新加入样本对已学完的样本有影响等，所以对 BP 算法有一些改进算法，如 Vogl 快速算法，二阶 BP 算法，二阶快速 BP 算法等等.

18.5.4　反馈型神经网络——Hopfield 模型

反馈型神经网络有离散型和连续系统两大类，主要取决于是用微分方程还是用差分模型描述. 它的一般结构如图 18－5所示.

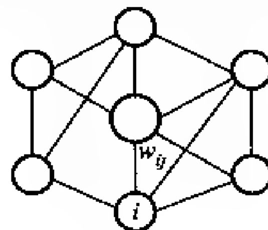


图 18－5

Hopfield 模型是典型的反馈型神经网络，是 J. J. Hopfield 和 D. W. Tank 在 1985 年提出的. Hopfield 网络可以用来解决联想记忆和约束优化问题的求解. 他们通过构造描述某一给定组合优化问题的能量函数，并反过来由能量函数构造出一个 ANN 模型，找到了获得 NP 问题近似最优化的方法.

Hopfield 模型神经网络是多个神经元相互组合的网络，并且神经元之间的相互结合强度是对称的，即 $w_{ij} = w_{ji}$，当 $w_{ii} = 0$ 时，称之为无自反馈的 Hopfield 网络，否则，称之为自反馈的 Hopfield 网络.

Hopfield 网络有异步和同步两种工作方式. 若在任一时刻 t，只有某一个神经元发生变化，而其余神经元的状态保持不变，称为异步方式；若有部分或全部神经元发生变化，称为同步方式.

由 N 个神经元构成的神经网络，第 i 个神经元在 t 时刻所接受的其他 $N-1$ 个神经元输入信号的总和 $u_i(t)$ 为

$$u_i(t) = \sum_{j \neq i} w_{ij} u_j(t) - \theta_i.$$

用符号函数作用于 $u_i(t)$，得到 $t+1$ 时刻的第 i 个神经元的输出值 $u_i(t+1)$. 故当某神经元输入端接受的信号强度超过 $u_i(t)$ 的某个阈值 θ_i 时，该神经元成兴奋状态.

下面给出 Hopfield 模型的算法.

算法 18.11

① 在网络中随机选择一个神经元；

② 求所选神经元 i 的输入总和：

$$u_i(t) = \sum_{j \neq i} w_{ij} u_j(t) - \theta_i;$$

③ 检查 $u_i(t) \geqslant 0$？

是，则 $u_i(t+1) = 1$，

否，则 $u_i(t+1) = 0$；

④ 神经元 i 以外的神经元 j 状态不变：

$$u_j(t+1) = u_j(t);$$

⑤ 转①.

附录Ⅲ　Matlab 及其应用

1.1　Matlab 简介

1.1.1　数学软件

熟悉计算机的人会接触各式各样的软件,诸如系统软件、编译软件、应用软件、杀毒软件等等. 20 世纪 80 年代以来,一类新的计算机软件——数学软件问世,即所谓 3M 软件.

1980 年,Wolfram 公司推出 Mathematica 软件;1982 年,Mathworks 公司推出 Matlab 软件;此后 Materloo 大学又推出一套教学软件,即 Maple 软件.

上述三种软件成为当今世界上最流行的数学软件,它们为解决现代科学计算、工程技术问题、生产经营管理问题提供了重要工具.

Mathematica 软件着重符号运算,作图功能;Matlab 软件着重数值计算,模拟功能;Maple 软件则是强大的教学软件.

1.1.2　什么是 Matlab

Matlab 的直接含义是矩阵实验室(Matrix laboratory). 最初,它是为两个矩阵软件包(Linpack,Eispack)提供接口技术,以后逐渐发展成为通用科技计算、视图交互系统和程序语言.

Matlab 语言区别于其他高级语言(如 Fortran 语言,C 语言)的最大特点是它的编辑、编译、连接和执行等四个步骤融为一体. 它能在同一画面上进行灵活操作,快速排除输入程序中的书写、语法乃至语意方面的错误,从而加快了用户编写、修改和调试程序的速度.

与其他高级语言比较,Matlab 的语法规则简单,更加贴近人的思维方式. 用 Matlab 写程序,犹如在一张演算纸上排列公式和求解问题,编程效率很高. 因此有人称它为"演算纸式"科学工程的算法语言. 下面我们举例说明 Matlab 的一些基本特点.

(1)基本数据单位是矩阵或数组,且无须定义矩阵或数组的维数,直接输入.

例 1 输入两个三维矩阵 $\boldsymbol{A},\boldsymbol{B}$,在 Matlab 的命令空间(Command)发布命令(指令)

```
a=[1  2  3;4  5  6;7  8  9];
b=[1  4  7;8  9  10;11  12  13];
```

也可写成

```
a=[1  2  3
   4  5  6
   7  8  9];
b=[1   4   7
   8   9   10
   11  12  13];
```

注 若在上述命令中不加分号而直接键入回车换行键,则屏幕显示:

```
a =
    1  2  3
    4  5  6
    7  8  9
b =
    1   4   7
    8   9   10
    11  12  13
```

若加分号再键入回车换行键,则此时光标转入第二行.

(2) 指令表达与数学习惯和工程中的常用习惯相似.

例 2 矩阵的加、减、乘.

在命令空间发布指令:

```
a = [1  2  3;4  5  6;7  8  9];
b = [1  4  7;8  9  10;11  12  13];
c = a + b
```

屏幕显示:

```
c =
    2   6   10
    12  14  16
    18  20  22
```

在命令空间发布指令:

```
x = [2  3  4  5;1  2  2  1];
y = [0  1  1;1  1  0;0  0  1;1  0  0];
z = x * y
```

屏幕显示：

```
z =

    8  5  6
    3  3  3
```

例 3　在 Matlab 中，矩阵的除法用于解线性方程组．

设 $\boldsymbol{Ax}=\boldsymbol{b}$，其中 $\boldsymbol{A}$ 为方阵，$\boldsymbol{x}$ 为列向量，$\boldsymbol{b}$ 为列向量，则指令为

x = a\b = inv(a) * b　　（即 $a^{-1}b$）

设 $\boldsymbol{xA}=\boldsymbol{b}$，其中 $\boldsymbol{x}$，$\boldsymbol{b}$ 均为行向量，则指令为

x = b/a = b * inv(a)　　（即 ba^{-1}）

注　在指令 a\b 中，a 为方阵，b 为列向量（a，b 同行数）．若输入时 b 为行向量，则应写成 a\b′．

(3) 易于扩展，允许用户自定义 M 文件，从而扩大内部函数的功能．

Matlab 有大量的内部函数，可以在命令空间直接调用．为了扩大内部函数的功能，还允许用户自己定义 M 文件（包括函数文件和文本文件）．M 文件在编辑空间（Edit）编写，在命令空间调用．我们称自定义的 M 文件为外部函数．

例 4　设

$$f(x)=\frac{1}{(x-0.3)^2+0.01}+\frac{1}{(x-0.9)^2+0.04}-6$$

式中，$x\in[-1,2]$，取步长为 0.01，试计算 $f(x)$ 的值，并画出图形，在求 $f(x)$ 的极小点．

解　求解上述问题，可分述为如下三个步骤．

① 自定义 Matlab 的外部函数，命名为 f.m

在 Matlab 的编辑空间录入两行：

```
function y = f(x)
y = 1./((x-0.3).^2 +0.01) +1./((x-0.9).^2 +0.04) -6;
```

② 将上述函数文件 f.m 存盘．

有两种存盘方式，一种是将该文件存入 Matlab 的文件夹 bin 中，此时在存盘对话框中选取“m 文件格式”，以后调用该文件时，在命令空间直接发布命令：

```
f
```

另一种存盘方式是将文件存入其他文件夹中，此时在存盘对话框中

选取“所有文件格式”，以后调用该文件时要首先确定路径. 例如，若将文件 f. m 存入 C 盘中的文件夹 My Documents 中，即

```
C:\My Documents\f.m
```

要在命令空间调用 f. m，则需要 Matlab 的内部函数 path 显示当前 Matlab 的搜索路径，即在命令空间发布如下指令：

```
Path('C:\My Documents',path)
f
```

③ 存盘文件 f. m 后退出编辑空间，进入 Matlab 的命令空间，发布如下指令：

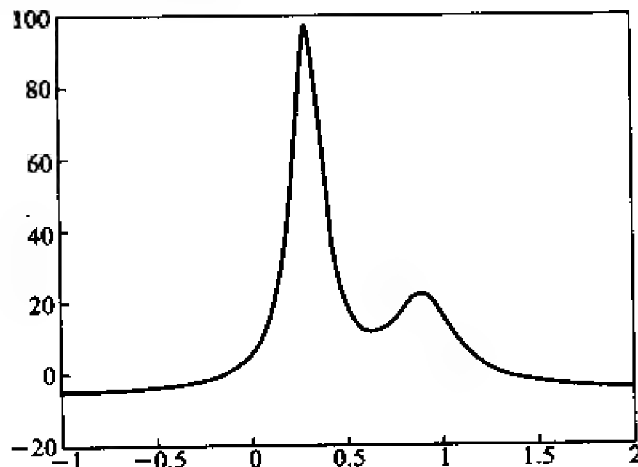
```
x = -1:0.01:2;
y = f(x)
```

屏幕显示出 $f(x)$ 的 300 个函数值.

若发布如下指令：

```
x = -1:0.01:2;
plot(x,f(x))
```

屏幕显示出 $f(x)$ 的图形(见图Ⅲ -1).



图Ⅲ -1

若发布命令：

```
xm = fmin('f',0.5,1)
```

屏幕显示 $f(x)$ 在 0.5 ~1 之间的一个极小点，即

```
xm =
    0.6370
```

当求其极小值时，可发布指令：

```
y = f(xm)
```

屏幕显示$f(x)$在极小点 xm 处的极小值：

```
y =
    11.252 8
```

上面我们用了三个主程序分别完成例 4 的三项要求（计算、画图、求极小点），其实我们可以用一个文本文件来完成这三项要求. 首先在编辑空间写入文本文件 L.m，它包含如下指令：

```
x = -1:0.01:2;
y = f(x);
plot(x,f(x));
xm = fmin('f',0.5,1)
```

将文本 L.m 在编辑空间存盘，然后在命令空间运行，即

```
L.m
```

屏幕将陆续显示$f(x)$的值、图形和一个极小点“xm”.

在上述指令中，$f(x)$是外部函数（自定义），它必须在编辑空间中编写存盘后方可调用；而 plot，fmin 等是内部函数，它可在命令空间直接调用.

（4）Matlab 易学易用

在前面的四个例子中，以下指令：

```
c = a + b,    z = x * y,    x = a\b,    xm = fmin('f',0.5,1)
```

与数学习惯十分相似，易学易用. 而这些指令若用其他高级语言来编写则相当困难，它们需用 Fortran 或 C 语言的几十条或几百条语句才能完成.

1.1.3　Matlab 的主要用途

Matlab 软件的主要用途如下：

（1）作为高等学校有关课程的基本教学工具. 例如数学实验、数学建模、实用线性代数、数值分析、最优化方法、最优控制、数理统计、数字信息处理、时间序列分析及动态系统仿真等课程均适宜使用 Matlab 作为教学工具. Matlab 软件是大学本科生、硕士生、博士生必须掌握的基本技能.

（2）在工程领域，Matlab 提供的工具箱以及由 M 文件扩展的工具箱成为工程师进行设计、科研的强大工具. Matlab 由主包（含几百个核心的内部函数）和 30 多个工具箱（Toolbox）构成，而 Toolbox 可分为两类：

功能性工具箱　扩充 Matlab 的符号计算功能、图视建模仿真功能、文字处理功能、硬件实时互换功能.

学科性功能箱　各种专业工具，例如，Optimal Toolbox，Control Toolbox，Communication Toolbox，Signal Toolbox 等等.

1.1.4 几点说明

(1) 在 Matlab 语言中,英文字母大小写要加以区别,如 A 与 a 代表不同的变量.

(2) 所有的内部函数、外部函数名均由小写字母构成,如 inv(a)有意义,但 INV(a)无意义.

(3) 标识符(变量,函数)以字母打头,最多 19 个字符.

(4) help(内部命令)提供在线帮助,如调用 helpfmin 将显示内部函数 fmin 的内容.

(5) demo(内部函数)提供功能演示,在 Matlab 的 Command 空间中有一个菜单项 Help,在其下拉菜单中有一个子菜单 Examples and Dermos,其中列出了最优化方法的各种算例.

(6) path(内部函数)显示当前 Matlab 搜索路径.

1.1.5 矩阵运算

Matlab 在进行矩阵运算方面,显得特别简捷、高效和方便

例 5 矩阵函数运算

输入矩阵

```
a=[1.5  1.1  1.3;2.1  2.3  2.5;  3.3  3.5  3.1]
```

屏幕显示:

```
a =

    1.5000    1.1000    1.3000

    2.1000    2.3000    2.5000

    3.3000    3.5000    3.1000
```

求 A 的转置.发布指令:

```
  a'
ans =
    1.5000    2.1000    3.3000
    1.1000    2.3000    3.5000
    1.3000    2.5000    3.1000
```

注 第一个指令中输入矩阵指定了赋值变量 a,其后未加分号“;”,故键入回车符后直接显示变量名 a 的结果;第二个指令中,转置矩阵未指定赋值变量,故键入回车符后显示变量名“ans”的结果.

求 $\boldsymbol{A}$ 的逆,发布指令:

```
inv(a)
ans =
     1.9565   -1.3768    0.2899
    -2.1014   -0.4348    1.2319
     0.2899    1.9565   -1.3768
```

求 $\boldsymbol{A}$ 的特征值,发布指令:

```
eig(a)
ans =
   6.9000
   0.3464
  -0.3464
```

对 $\boldsymbol{A}$ 作 $\boldsymbol{LU}$ 分解,发布指令:

```
[L,U] = lu(a)
L =
   0.4545    1.0000         0
   0.6364   -0.1481    1.0000
   1.0000         0         0
U =
   3.3000    3.5000    3.1000
        0   -0.4909   -0.1091
        0         0    0.5111
```

关于矩阵函数运算的内容还很多,如对矩阵进行正交变换、奇异值分解、求矩阵的特征值和特征向量、求矩阵的和条件数以及建立矩阵函数,等等.

例 6　矩阵的截取与合成

输入矩阵　a = [1　2　3;4　5　6;7　8　9;10　11　12];

指令 a(i,:)表示取第 i 行所有的元素;a(:,i)表示取 $\boldsymbol{A}$ 的第 i 列所有的元素. 例如发布指令:

```
a(4,:)            (取 A 的第 4 行)
ans =
   10    11    12
a(1:3,:)          (取 A 的第 1～3 行)
ans =
    1     2     3
    4     5     6
```

```
    7     8     9
a(:,1:2)          (取 A 的第 1～2 列)
ans =
    1     2
    4     5
    7     8
   10    11
a(:,3)          (取 A 的第 3 列)
ans =
    3
    6
    9
   12
```

综上,a(i,:)表示取第 i 行所有的元素,a(:,j)表示取第 j 列所有的元素,a(i:j,:)表示取第 i 行至第 j 行所有的元素,a(:,i:j)表示取第 i 列至第 j 列所有的元素.

设 $\boldsymbol{A}$ 是例 6 中的输入矩阵,再输入 $\boldsymbol{r}$,即

```
r=[13  14  15];
```

发布指令:

```
[a;r]
```

即将 $\boldsymbol{r}$ 添入到 $\boldsymbol{A}$ 的最后一行合成新的矩阵,屏幕显示:

```
ans =
    1    2    3
    4    5    6
    7    8    9
   10   11   12
   13   14   15
```

也可以指定赋值变量 $\boldsymbol{b}$,即

```
b=[a;[13  14  15]]
b =
    1    2    3
    4    5    6
    7    8    9
   10   11   12
   13   14   15
```

例 7　矩阵的除法(左除"\"与右除"/").

设 A 是 $m\times m$ 阶方阵,B 是 $m\times n$ 阶矩阵且 $\det(A)\neq 0$,其命令为

A\B ⇔A * Z = B　　　(Z 是 $m\times n$ 阶矩阵,等价于解矩阵方程)

⇔A * Z(:,j) = B(:,j)($j=1,\cdots,n$;等价于解 n 个线性方程组)

⇔Z(:,j) = A\B(:,j) = inv(A) * B(:,j)

($j=1,\cdots,n$;求得 n 个方程组的解)

A/B = (A\B)'

例 8　矩阵的除法(点左除".\"与点右除"./")

设 A、B 均为 $m\times n$ 阶矩阵,"A.\B"表示 B 的所有元素分别被 A 的所有对应元素相除后所形成的 $m\times n$ 阶矩阵;"A./B"表示 A 的所有元素分别被 B 的所有对应元素相除后所形成的 $m\times n$ 阶矩阵.

输入以下指令:

```
A = [1  2  3];
B = [4  5  6];
C = A.\B
```

屏幕显示:

```
C =
    4.0000    2.5000    2.0000
D = A./B
```

屏幕显示:

```
D =
    0.2500    0.4000    0.5000
```

例 9　特殊矩阵(友矩阵、零矩阵、常值矩阵)

友矩阵(compan)

设三次多项式为 x^3-7x+6,欲求该多项式的零点,即求 $x^3-7x+6=0$ 的三个根.

解　输入

```
p = [1  0  -7  6];
a = compan(p)
```

屏幕显示:

```
a =
    0    7   -6
    1    0    0
    0    1    0
```

再输入

```
eig(a)
ans =
      -3.0000
       2.0000
       1.0000
```

友矩阵 **A** 的三个特征值即为所给三次多项式的三个零点.

零矩阵(zeros):

zeros(i,j)表示产生 $i \times j$ 阶的零矩阵;zeros(size(a))表示产生与矩阵 **A** 同阶的零矩阵.

常数矩阵(ones):

ones(i,j)表示产生 $i \times j$ 阶的元素全为 1 的矩阵;ones(size(a))表示产生与 **A** 同阶的元素全为 1 的矩阵.

1.2 最优化方法计算

1.2.1 无约束极值算例

1) 一元函数求极小点

内部命令 fmin('f',a,b)表示求一元函数 $f(x)$ 在区间[a,b]内的极小点,其中 f 是在 Edit 空间编辑的一个外部函数(M 文件).

内部命令 fminsearch('f',x)表示求 $f(x)$ 在初始点 x 邻近的极小点.

例 1 $f(x)=(x-3)\sqrt{x}, x_0=0.5$.

解 (1) 在 Edit 空间编辑一个外部函数,取名"fun. m"

```
function f = fun(x)
f = (x - 3) * sqrt(x);
```

(2) 将 M 文件 fun 存盘.

(3) 在 Command 空间调用 fun.

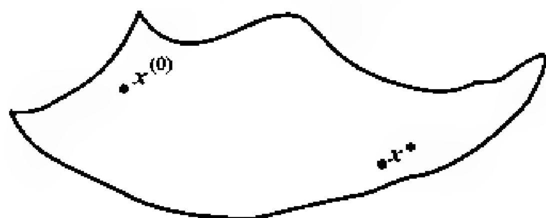
```
x0 = 0.5;
x = fminsearch('fun',x0);
x =
     1.0000            (f(x)在 x=0.5 附近的极小点)
fun(x)
ans =
      -2.0000          (f(x)在极小点所对应的函数值)
```

2) 多元函数求极值

内部命令 fminu('f',x)表示求多元函数$f(\boldsymbol{x})$在向量$\boldsymbol{x}$邻近的局部极小值.

例 2　香蕉函数(Banana Function)

设$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$,其图形如图Ⅲ-2所示.$\boldsymbol{x}^{(0)} = (-1.9, 2)^T$为初始点,$\boldsymbol{x}^* = (1,1)^T$为极小点,$f(\boldsymbol{x}^*) = 0$



图Ⅲ-2

演示程序　在 Matlab 的 Command 空间中,选择菜单项 Help 的子菜单 Demos,即列出求解该香蕉函数的六种算法的演示图形.这些无约束优化算法分别是:Steepest 算法,Simplex 算法,DFP 算法,BFGS 算法,Gauss-Newton 算法,Levenbrg-Marguardt 算法(简称 L-M 算法).

解　(1) 自定义外部函数,建立 M 文件.

```
function y = fun1(x)
y = 100 * (x(2) - x(1)^2)^2 + (1 - x(1))^2;
```

(2) 存盘 fun1.m.

(3) 调用 fun1:

```
x = [-1.9,2];
x = fminu('fun1',x)
```

屏幕显示:

```
x =
    1.0000    1.0000
fun1(x)
ans =
    1.1730e-010
```

事实上,Matlab 在演示中提到的六种无约束算法其效果是不一样的.通过演示程序可以看出:从初始点$\boldsymbol{x}^{(0)}$出发能全程达到极小点$\boldsymbol{x}^*$的算法有单纯形法(Simplex)、DFP 算法、BFGS 算法以及 L-M 算法;从$\boldsymbol{x}^{(0)}$出发只能搜索到中途的算法是最快速下降法(Steepest);只能从$\boldsymbol{x}^*$的邻近点出发而达到$\boldsymbol{x}^*$的算法是 Gauss-Newton 算法(G-N);而从$\boldsymbol{x}^{(0)}$出发最快速达到$\boldsymbol{x}^*$的算法有 DFP,BFGS 算法.

1.2.2　约束极值

1）线性规划算法

$$(LP)\ \min\ \ \boldsymbol{c}^{T}\boldsymbol{x}$$
$$\text{s.t.}\begin{cases}\boldsymbol{A}\boldsymbol{x}\leqslant \boldsymbol{b},\\ \boldsymbol{x}\geqslant 0.\end{cases}$$

其中 $\boldsymbol{A}_{m\times n},\boldsymbol{x}\in \mathbf{R}^{n},\boldsymbol{c}\in \mathbf{R}^{n},\boldsymbol{b}\in \mathbf{R}^{m}$.

内部命令 linprog(f,A,b)表示求上述线性规划的最优解，其中 f 为线性规划的目标函数的价值系数

$$f=\boldsymbol{c}^{T}\boldsymbol{x}=c_1x_1+c_2x_2+\cdots+c_nx_n$$

例 3

$$\min(2x-3y)$$
$$\text{s.t.}\begin{cases}x+y<10\\ -x+y<-2\\ x>1(-x<-1)\end{cases}$$

解　(1) 自定义外部函数，建立 M 文件 fun2.m(用 x 表示二维向量，其分量 x(1),x(2)分别表示题中的 x,y)，即

```
function f = fun2(x)
f = 2 * x(1) - 3 * x(2);
```

(2) 存盘 fun2.m

(3) 调用：

```
A = [1  1; -1  1; -1  0];
b = [10  -2  -1];
f = [2  -3];
x = linprog(f,A,b)
x =
     6.0000
     4.0000
fun2(x)
ans =
     5.6843e-014
```

2）非线性规划算法(基本算法)

$$(NLP)\qquad\begin{cases}\min f(\boldsymbol{x}),\\ \text{s.t. } g(\boldsymbol{x})\leqslant 0.\end{cases}$$

其中 $f(\boldsymbol{x})$ 为 n 元非线性函数，$\boldsymbol{g}(\boldsymbol{x})$ 为向量值函数，$\boldsymbol{g}(\boldsymbol{x})=(g_1(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x}))^T$，要求 $f(x)$ 与 $g_i(\boldsymbol{x})(i=1,\cdots,m)$ 其中至少有一个为非线性函数.

内部命令 constr('fg',x)，f 代表目标函数，g 代表约束函数，x 代表初始点.

例 4

$$\min f(\boldsymbol{x}) = e^{x_1}(4x_1^2+2x_2^2+4x_1x_2+2x_2+1)$$

$$\text{s.t.}\begin{cases}1.5+x_1x_2-x_1-x_2\leqslant 0,\\ -x_1x_2-10\leqslant 0.\end{cases}$$

解　(1) 自定义外部函数 fun3.m

```
function[f,g] = fun3(x)
f = exp(x(1)) * (4 * x(1)^2 +2 * x(2)^2 +4 * x(1) * x(2) +
2 * x(2) +1);
g(1) =1.5 + x(1) * x(2) - x(1) - x(2);
g(2) = - x(1) * x(2) - 10;
```

(2) 存盘 fun3.m

(3) 调用：

```
x = [ -1,1];
x = constr('fun3',x)
x =
    -9.5474    1.0474
[f,g] = fun3(x)
f =
    0.0236
g =
    1.0e -014 * 0.1110    -0.1776
```

3) 非线性规划算法(带有参数的情形)

例 5　下界条件与上界条件

在例 4 中，若对决策变量 $\boldsymbol{x}$ 限定在某一区域内，如 $\boldsymbol{vlb}<\boldsymbol{x}<\boldsymbol{vub}$，其中 $\boldsymbol{vlb},\boldsymbol{vub}$ 均为二维向量，则可用如下语句：

```
x = constr('fun3',x,options,vlb,vub);
```

又如在例 4 中，要求 $\boldsymbol{x}>0$，则调用时应发布如下指令：

```
x = [ -1   1];          (赋初值)
options = [        ];   (阐述说明向量使用缺省选项)
vlb = [0   0];          (赋下界)
```

```
vub = [      ];          (无上界)
x = constr('fun3',x,options,vlb,vub);
```

经 7 次迭代后,问题的解为

```
x =
   0      1.5000
[f,g] = fun3(x)
f =
   8.5 000
g =
   0    -10
```

若决策变量 $\boldsymbol{x}$ 的一部分分量有上、下界的限制,则可通过对 ***vlb***,***vub*** 向量部分赋值来实现,当 ***vlb***,***vub*** 的分量数目 t 比向量 $\boldsymbol{x}$ 的分量数目 n 少时,则 $\boldsymbol{x}$ 中只有前 t 个分量被限定有界.

例 6　不等式、等式混合约束

例 4 中,若再加入等式约束 $x_1 + x_2 = 0$,则相应的程序变动如下.

解　(1) 自定义外部函数

```
function[f,g] = fun4(x)
f = exp(x(1)) * (4 * x(1)^2 + 2 * x(2)^2 + 4 * x(1) * x(2) + 2
* x(2) + 1);
g(1) = x(1) + x(2);
g(2) = 1.5 + x(1) * x(2) - x(1) - x(2);
g(3) = - x(1) * x(2) - 10;
```

(2) 存盘 fun4.m

(3) 调用:

```
x = [ -1   1];
options(13) = 1;
x = constr('fun4',x,options)
```

经 11 次迭代后的解:

```
x =
   -1.2247      1.2247
[f,g] = fun4(x)
f =
   1.8951
g =
   0.0000    -0.0000    -8.5000
```

注 参数说明向量 options 共有 18 个分量,包含了在优化程序中需要用到的所有参数.第 13 个分量的功能是等式约束条件.例如 options(13) = 1 表明第一个约束条件是等式约束,options(13) = 3 表明第 1 至第 3 个约束条件是等式约束,因此 1,3 表示等式约束的个数.等式约束条件必须放置在变量 g 的前几个分量中.

1.2.3 线性最小二乘问题

众所周知,已知实测数据(t_i, y_i), $i = 1,2,\cdots,m$,其中 $t_i \neq t_j (i \neq j)$, $y_i = f(t_i)$, $i = 1,2,\cdots,m$,若用多项式

$$y = \sum_{j=0}^{n} a_j t^j, n \ll m$$

去拟合曲线 $y = f(t)$,则问题归结为求超定方程组的解,即解方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$,其中

$$\boldsymbol{A} = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^n \\ 1 & t_2 & t_2^2 & \cdots & t_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_m & t_m^2 & \cdots & t_m^n \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

可以证明 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的非常义解(最小二乘解)即是如下线性最小二乘问题的解:

$$\min_{x \in \mathbb{R}^{n+1}} \| \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \|_2^2.$$

Matlab 给出了非负的线性最小二乘问题的解法:

$$\begin{cases} \min\limits_{x \in \mathbb{R}^{n+1}} \| \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \|_2^2, \\ \text{s.t.}\ \boldsymbol{x} \geqslant 0. \end{cases}$$

Matlab 的内部函数为 lsqnonneg(A,b),其中 A 表示超定方程的系数矩阵,b 表示右端项.

例 7 解超定方程组

$$\begin{cases} 2x + 4y = 11, \\ 3x - 5y = 3, \\ x + 2y = 6, \\ 2x + y = 7. \end{cases}$$

上述方程组得矩阵形式为 $\boldsymbol{A}\overline{\boldsymbol{X}} = \boldsymbol{b}$,其中

$$\boldsymbol{A} = \begin{bmatrix} 2 & 4 \\ 3 & -5 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} x \\ y \end{bmatrix}, \boldsymbol{b} = [11, \ 3, \ 6, \ 7]^{\mathrm{T}}$$

解 直接对 Matlab 的 Command 空间发布指令：

```
A = [2  4;3  -5;1 2;2 1];
b = [11 3 6 7];
X = lsqnonneg(A,b')
X =
    3.0403
    1.2418
```

例 8 已知实测数据如下：

x_i	-1.00	-0.75	-0.50	-0.25	0	0.25	0.50	0.75	1.00
y_i	0.2209	0.3295	0.8826	1.4392	2.003	2.5645	3.1334	3.7061	4.2836

试用一次、二次多项式拟合 $y=f(x)$.

解 (1) 一次多项式拟合. 设

$$y = a + bx,$$

并将实测数据 $(x_i, y_i)_{i=1,\cdots,9}$ 代入上式得超定方程组

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y},$$

其中：

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1.0000 \\ 1 & -0.7500 \\ 1 & -0.5000 \\ \vdots & \vdots \\ 1 & 1.0000 \end{bmatrix}_{9\times 2}, \boldsymbol{X} = \begin{bmatrix} a \\ b \end{bmatrix}, \boldsymbol{Y} = \begin{bmatrix} 0.2209 \\ 0.3295 \\ \vdots \\ 4.2836 \end{bmatrix}_{9\times 1}.$$

Matlab 指令：

```
A = [1 -1.0000;1  -0.7500;1  -0.5000;1  -0.2500;
     1  0.0000;  …;1  0.2500;1  0.5000;1  0.7500;
     1  1.0000];
Y = [0.2209  0.3295  0.8826  1.4392  2.003
     2.5645  …  3.1334  3.7061  4.2836];
X = lsqnonneg(A,Y')
X =
     2.0625
     2.1338
```

(2) 二次多项式拟合，设

$$y = a_0 + a_1 x + a_2 x^2,$$

将实测数据 $(x_i, y_i)_{i=1,\cdots,9}$ 代入上式得超定方程组

$$AX = Y,$$

其中

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_9 & x_9^2 \end{bmatrix}, X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_9 \end{bmatrix}.$$

Matlab 指令：

```
A = [1   -1.0000   1.0000;1   -0.7500   0.5625;…;
     1   -0.5000   0.2500;…;1   -0.2500   0.0625;
     1   0.0000   0.0000;1   0.2500   0.0625;…;
     1   0.5000   0.2500;1   0.7500   0.5625;1   1.0000
     1.0000];
Y = [0.2209   0.3295   0.8826   1.4392   2.0003
     2.5645   …   3.1334   3.7061   4.2836];
X = lsqnonneg(A,Y')
X =
     1.9599
     2.1338
     0.2455
```

注　对于三次多项式拟合，使用 Matlab 命令将导致失败. 因为三次多项式拟合时，其系数向量$(a_0,a_1,a_2,a_3)^T=(1.9599,2.4272,0.2455,-0.3881)^T$中的第 4 个分量为负数，这就违反了非负的条件. 在本附录 1.3.4 中我们将介绍多项式拟合的一般方法（去掉 x 非负的限制）.

1.2.4　非线性最小二乘问题

数学模型：$\min\limits_{x\in\mathbf{R}^n} \| F(x) \|_2^2$ 其中

$$F(x) = (f_1(x), f_2(x), \cdots, f_m(x))^T, x = (x_1, x_2, \cdots, x_n)^T.$$

通常将其写成平方和的极小形式

$$\min_{x\in\mathbf{R}^n} S(x) = \sum_{i=1}^{m} f_i^2(x)$$

上述问题的实际意义为：给定实测数据$(t_i, y_i)_{i=1,\cdots,m}$，其中$t_i \neq t_j (i\neq j)$，$y_i = f(t_i)$，如果拟合函数不是代数多项式形式，如

$$y(t,x) = x_1 + x_2\exp(tx_3)$$

（一元三参数问题，t 为实数，x 为三维向量），

$$y(t,x) = \frac{x_1x_3t_1}{1 + x_1t_1 + x_2t_2}$$

（二元三参数问题，t 为二维向量，x 为三维向量），

则采用非线性最小二乘法作偏差平方和函数，即

$$S(x) = \sum_{i=1}^{m}[y(t_i,x) - y_i]^2 = \sum_{i=1}^{m}f_i^2(x),$$

其中偏差为 $y(t_i,x) - y_i = f_i(x), i = 1,\cdots,m$

Matlab 函数 lsqnonlin('f ',x) 给出了求解上述非线性最小二乘问题的算法，其中"f"代表平方和函数 $S(x)$，"x"是给定的初始点.

例 9　一元三参数问题，设已知数据如下：

t_i	$t_1 = -5$	$t_2 = -3$	$t_3 = 1$	$t_4 = 3$	$t_5 = 5$	$t_6 = -1$
y_i	$y_1 = 127$	$y_2 = 151$	$y_3 = 421$	$y_4 = 460$	$y_5 = 426$	$y_6 = 379$

拟合函数 $y(t,x) = x_1 + x_2\exp(tx_3)$，

$$S(x) = \sum_{i=1}^{6}[y(t_i,x) - y_i]^2 = \sum_{i=1}^{6}[x_1 + x_2\exp(t_ix_3) - y_i]^2$$

给定 $x^{(0)} = (580, -180, 0.16)^T, S^{(0)} = 0.3 \times 10^5$.

解　(1) 自定义外部函数.

```
function S = f(x)
t = [ -5   -3   1   3   5   -1];
y = [127   151   421   460   426   379];
S = 0.0;
for i = 1:6
S = S + (x(1) + x(2) * exp(t(i) * x(3)) - y(i))^2
end
```

(2) 存盘 f. m

(3) 调用：

```
x = [580    - 180   0.16];
x = lsqnonlin('f ',x)
x =
      592.3862    -190.5805    -0.1885
```

```
S = f(x)
S =
      2.1310e+0.04
```

1.3 数据分析

1.3.1 数据的输入和输出

1）键盘输入单个参数（input 命令）

例 1 $f(x)=[(x-0.3)^2+0.01]+[(x-0.9)^2+0.04]-6$，计算$f(x)$的值，其中 x 是待输入参数.

解 （1）建立函数文件 f.m（存盘）.

```
function y = f(x)
y = ((x-0.3)^2+0.01)+((x-0.9)^2+0.04)-6;
```

（2）建立文本文件 s.m（存盘）.

```
x = input('input x:');
y = f(x)
```

（3）运行 s.m

```
s            （在 Command 空间发布）
```

屏幕显示

```
input x:
     5           （Command 空间发布）
y =
     32.9500
```

2）矩阵输入（小型矩阵）

例 2 解线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$，其中

$$\boldsymbol{A}=\begin{bmatrix}1 & 1/2 & 1/3\\ 1/2 & 1/3 & 1/4\\ 1/3 & 1/4 & 1/5\end{bmatrix},\boldsymbol{x}=(x_1,x_2,x_3)^{\mathrm{T}},$$

$$\boldsymbol{b}=(11/6,13/12,47/60)^{\mathrm{T}}.$$

解 （1）建立文本文件 solution.m（存盘）.

```
a = [1   1/2   1/3   ;1/2   1/3   1/4;1/3   1/4   1/5];
b = [11/6   13/12   47/60];
x = a\b'
```

（2）运行 solution.m

solution （在 Command 空间发布）

屏幕显示

x =

1.0000

1.0000

1.0000

3）用数据文件输入（大型矩阵）

例 3 经济（longley）数据

某国（或某地区）1947 年—1962 年（共 16 年）的经济数据如表Ⅲ－1 所示

表Ⅲ－1

年份	GNP	GDP	失业率	军队人数	人口	就业率
1947	×	×	×	×	×	×
1948	×	×	×	×	×	×
1949	×	×	×	×	×	×
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1962	×	×	×	×	×	×

注：GNP 表示紧缩率，GDP 表示国民生产总值

附表 1－1 内的数据构成 16×7 阶矩阵，若用前面两种方式输入显然不宜，下面用数据文件输入．

解 （1）建立数据文件（或文本文件）longley. dat，在 Edit 环境下写入如下数据块（16 行，7 列）：

× × × × × × ×

⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮

× × × × × × ×

（2）存盘 longley. dat.

（3）调用：

longley

屏幕显示 16×7 数据块，再调用：

[n,p] = size(longley)

屏幕显示：

n =

16

```
p =
7
```

(4) 也可建立文本文件 longley. m. 设赋值变量名为 idata，在 Command 空间键入

```
idata = [ ×  ×  ×  ×  ×  ×  × ;
      ……  ×  ×  ×  ×  ×  ×  × ]
[n,p] = size(idata)
```

建立文本文件 longley. m 存盘后再调用：

```
longley. m
```

屏幕分别显示出 16 ×7 阶的以变量 idata 标志的数据块以及与上面相同的 n，p 值.

1.3.2 列数据分析

下列内部函数具有列数据的分析功能：

给定 longley 16 ×7 阶矩阵

max(longley)	longley 的各列的最大值所构成的行向量
min(longley)	longley 的各列的最小值所构成的行向量
mean(longley)	longley 的各列的平均值所构成的行向量
mediam(longley)	longley 的各列的中值所构成的行向量
std(longley)	longley 的各列的标准差(均方差)所构成的行向量
sum(longley)	longley 的各列元素总和所构成的行向量
prod(longley)	longley 的各列元素乘积所构成的行向量

若给定 x 为列向量，则 max(x)，min(x)，…，prod(x) 均为数；若 x 为行向量，则它们仍为行向量.

例 4 求矩阵 longley 的误差矩阵.

一个给定矩阵的误差矩阵是由该矩阵每一列元素分别减去该列元素的均值所形成的.

解 发布如下指令：

```
[n,p] = size(longley)
```

屏幕显示：

```
n =
    16
p =
    7
e = ones(n,1)
```

屏幕显示：

$$e=\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}$$

发布指令

m = mean(longley)

屏幕显示：

m =

1.0e+0.03

1.9550　1.0120　0.3975　0.3351　0.2798　0.1174

0.0660

发布指令

mean = Longley - e * m

屏幕显示 longley 的误差矩阵(16×7 阶).

注　设 $\boldsymbol{m}=[m(1)\ m(2)\ m(3)\ m(4)\ m(5)\ m(6)\ m(7)]$.

$$\boldsymbol{e}\cdot\boldsymbol{m}=\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}[m(1)\quad m(2)\quad\cdots\quad m(7)]$$

$$=\begin{bmatrix}m(1) & m(2) & \cdots & m(7)\\ m(1) & m(2) & \cdots & m(7)\\ \vdots & \vdots & & \vdots\\ m(1) & m(2) & \cdots & m(7)\end{bmatrix}.$$

Longley - e * m 是从 longley 的各列分别减去各列的均值 $m(i)(i=1,\cdots,7)$而形成的误差矩阵，即

$$\begin{bmatrix}\times-m(1) & \times-m(2) & \cdots & \times-m(7)\\ \times-m(1) & \times-m(2) & \cdots & \times-m(7)\\ \vdots & \vdots & & \vdots\\ \times-m(1) & \times-m(2) & \cdots & \times-m(7)\end{bmatrix}.$$

1.3.3　实测数据归一化(标准化)

当试图对某组数据进行拟合或回归时，首先应对数据标准化，以改善结果的精确度.

设数据块 longley 为 16 ×7 阶矩阵. 标准化该数据的方法为:

(1) 移去均值,得误差矩阵.

```
e = ones(length(longley),1);
x = longley - e * mean(longley);
```

(2) 标准化,即将 x 化为单位标准差矩阵.

```
std(x);
e * std(x);
x = x./(e * std(x));
```

注 设 $\mathrm{std}(x) = [S_1, S_2, \cdots, S_7]$,则

$$e.\,\mathrm{std}(x) = \begin{bmatrix} S_1 & S_2 & \cdots & S_7 \\ S_1 & S_2 & \cdots & S_7 \\ \vdots & \vdots & & \vdots \\ S_1 & S_2 & \cdots & S_7 \end{bmatrix}$$

(3) 对单位标准差矩阵,计算相关系数矩阵.

```
x' * x/(n - 1)          (n = length(longley))
ans =
        ×   ×   ×   ×   ×   ×   ×
        ×   ×   ×   ×   ×   ×   ×                    (16 ×7 阶)
        ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮
        ×   ×   ×   ×   ×   ×   ×
```

其中相关系数 × ≈1.

先不对 longley 标准化,直接求相关系数:

```
corrcoef(longley)
ans =
        ×   ×   ×   ×   ×   ×   ×
        ×   ×   ×   ×   ×   ×   ×                    (16 ×7 阶)
        ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮
        ×   ×   ×   ×   ×   ×   ×
```

其中相关系数"×"最小值为 -0.1774,最大值为 1.0000.

1.3.4 多项式拟合

下面介绍多项式拟合的两种方法.

1) 解超定方程组 $Ax = b$(对 x 不作非负的限制)

设实测数据 $(t_i, y_i)_{i=1,\cdots,m}$,令拟合多项式

$$y(t) = a_0t^n + a_1t^{n-1} + \cdots + a_{n-1}t + a_n$$

将(t_i, y_i)代入上式，得

$$y(t_i) = a_0t_i^n + a_1t_i^{n-1} + \cdots + a_{n-1}t_i + a_n \overset{令}{=} y_i, i = 1,2,\cdots,m;$$

从而形成超定方程组 $Ax = b$，其中

$$A = \begin{bmatrix} t_1^n & t_1^{n-1} & \cdots & t_1 & 1 \\ t_2^n & t_2^{n-1} & \cdots & t_2 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ t_m^n & t_m^{n-1} & \cdots & t_m & 1 \end{bmatrix}_{m\times(n+1)},$$

$$x = (a_0, a_1, \cdots, a_{n-1}, a_n)^T,$$

$$b = (y_0, y_1, \cdots, y_{n-1}, y_n)^T.$$

该超定方程组得最小二乘解，即极小问题 $\min\limits_{x\in R^{n+1}} \| Ax - b \|_2^2$ 的解 x^* 为所求的拟合系数

$$x^* = (a_0^*, a_1^*, \cdots, a_{n-1}^*, a_n^*)^T.$$

实现上述过程的 Matlab 程序步骤为：

第一步，形成超定方程组的系数矩阵，即

```
for j = 1:(n+1)
A(:,j) = t.^(n+1-j);
end
```

第二步，解超定方程组 $Ax = b$，即

```
x = A\b
```

注 超定方程的系数矩阵 A 是超定矩阵，因而不能定义逆阵，但可定义广义逆阵. 这里取 A 的左广义逆阵 $A_左^+ = (A^{-1}A)^TA^T$，故 $x = A_左^+ b = (A^TA)^{-1}A^Tb$，即

```
x = inv(A'A) * A' * b;
```

设实测数据$(t_i, y_i)_{i=0,1,\cdots,10}$如下表：

t_i	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
y_i	2.1	2.3	2.5	2.9	3.2	3.3	3.8	4.1	4.9	5.4	5.8

用三次多项式拟合上述实测数据($n = 3$)

下面编写 Matlab 程序：

(1) 定义文本文件 fitting.m

```
t = 0:0.1:1;
b = [2.1 2.3 2.5 2.9 3.2 3.3 3.8  4.1 4.9 5.4 5.8];
```

```
n = input('input n:');
for j = 1:(n+1)
A(:,j) = t'.^(n+1-j);
end
x = A\b
```

(2) 存盘 fitting. m

(3) 运行文本文件:

```
fitting
```

屏幕显示:

```
input n: 3
```

屏幕显示:

```
x =
    0.6993
    1.2005
    1.8869
    2.1077
```

(4) 若需画出拟合函数的图形,只需在文本文件 fitting. m 中加入以下内容:

```
t=0.05:0.1:1.05;(0.05 为初值,0.1 为步长,1.05 为终值)
u=polyval(x,t);
plot(t,u,s,b,'+');
```

在上述各语句中,各标志变量的含义为:

x——实测数据自变元值; u——拟合多项式的值;

b——实测数据函数值; s——拟合系数向量;

t——等距原点.

(5) 运行

```
fitting
```

屏幕显示图形如图Ⅲ-3,图中实线代表拟合曲线,"+"代表实测数据的散点.

2) 直接调用 Matlab 内部函数

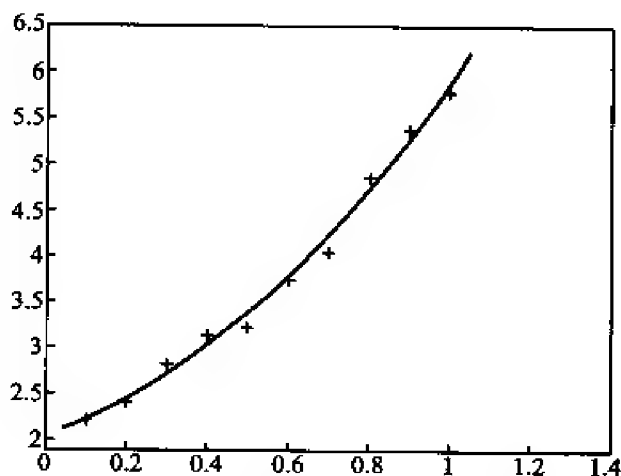
直接调用 Matlab 内部函数

```
polyfit(x,y,n)
```

其中参数向量 x,y 表示实测数据,n 表示拟合多项式的次数.

(1) 建立文本文件 fitting. m

```
t=0:0.1:1;
```



图Ⅲ－3

```
b = [2.1 2.3 2.5 2.9 3.2 3.3 3.8  4.1 4.9 5.4 5.8];
n = input('input n:');
x = polyfit(t,b,n)
```

（2）存盘 fitting.m

（3）运行：

```
fitting
```

屏幕显示：

```
input n:3
```

屏幕显示：

```
x =
    0.6993     1.2005     1.8869     2.1077
```

1.3.5　多元线性回归

为处理例 3 中的经济数据，现考虑六元线性回归问题. 设实测数据如下：

$$(t_{i1}, t_{i2}, t_{i3}, t_{i4}, t_{i5}, t_{i6}, y_i)\ (i = 1, \cdots, m),$$

其中回归自变元 $t_1, t_2, \cdots, t_6$ 以及回归因变元 y 均为随机变量. 设六元线性回归方程

$$\tilde{y} = \hat{a}_1 t_1 + \hat{a}_2 t_2 + \hat{a}_3 t_3 + \hat{a}_4 t_4 + \hat{a}_5 t_5 + \hat{a}_6 t_6 + \hat{a}_7,$$

即对于 $t_1, t_2, \cdots, t_6$ 的每一组值，有

$$y \sim N(a_1 t_1 + a_2 t_2 + \cdots + a_6 t_6 + a_7, \delta^2),$$

其中 $a_i\ (i = 1, 2, \cdots, 6)$，$\delta^2$ 均为未知参数，且不依赖 $t_i\ (i = 1, 2, \cdots, 6)$，亦即

$$y = a_1t_1 + a_2t_2 + \cdots + a_6t_6 + a_7 + \varepsilon, \quad \varepsilon \sim N(0,\delta^2)$$

下面利用所给实测数据去估计参数 $a_i(i=1,2,\cdots,6)$ 的值 $\hat{a}_i(i=1,2,\cdots,6)$，从而得到回归方程

$$\tilde{y} = \hat{a}_1t_1 + \hat{a}_2t_2 + \hat{a}_3t_3 + \hat{a}_4t_4 + \hat{a}_5t_5 + \hat{a}_6t_6 + \hat{a}_7$$

参数 $\hat{a}_i(i=1,2,\cdots,6)$ 的估计，采用极大似然法

记

$$\begin{aligned} \boldsymbol{t}_1 &= [t_{11} \quad t_{21} \quad \cdots \quad t_{m1}], \\ \boldsymbol{t}_2 &= [t_{12} \quad t_{22} \quad \cdots \quad t_{m2}], \\ &\cdots \quad \cdots \quad \cdots \quad \cdots \\ \boldsymbol{t}_6 &= [t_{16} \quad t_{26} \quad \cdots \quad t_{m6}], \\ \boldsymbol{y} &= [y_1 \quad y_2 \quad \cdots \quad y_m], \\ \boldsymbol{x} &= [\hat{a}_1 \quad \hat{a}_2 \quad \cdots \quad \hat{a}_6 \quad \hat{a}_7]. \end{aligned}$$

从而形成超定方程组的系数矩阵

$$\boldsymbol{A} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{16} & 1 \\ t_{21} & t_{22} & \cdots & t_{26} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{m6} & 1 \end{bmatrix}_{m\times 7}$$

由超定方程组 $\boldsymbol{Ax}=\boldsymbol{y}$ 解得

$$\boldsymbol{x} = \boldsymbol{A}/\boldsymbol{y}'$$

例 5 给出 16×7 的数据块 longley，现以 longley 的前 6 列数据（年份，GNP，GDP，失业率，军队人数，人口数）线性回归第 7 列数据——就业率.

解 （1）标准化 longley，得单位标准差矩阵.

```
x = x./(e * std(x))
```

（2）令：

```
y = x(:,7)
```

（3）形成超定方程组的系数矩阵.

```
A = [x(:,1:6)   ones(y)]
```

（4）形成超定方程组.

$$\boldsymbol{Ax} = \boldsymbol{y}$$

（5）求解：

$$\boldsymbol{x} = \boldsymbol{A}/\boldsymbol{y}'$$

注 $\boldsymbol{x} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_6, \hat{a}_7)^{\mathrm{T}}$ 代表回归方程的参数向量的估计值，即

为问题之所求.

下面用 Matlab 编程:

(1) 建立文本文件 coef. m

```
e = ones(length(longley),1);
x = longley - e * mean(longley);
x = x./(e * std(x));
y = x(:,7);
A = [x(:,1:6)   ones(y)];
x = A\y
```

(2) 存盘 coef. m

(3) 运行:

```
coef
```

屏幕显示:

```
x =
    1.0e+003 *
    0.0018
    0.0000
    0.0000
    0.0000
    0.0000
    0.0000
   -3.4982
```

约束最优化方法习题

第10、11章习题

1. 有下列线性规划问题：

(1)
$$\min z=4x_1+3x_2+8x_3,$$
$$\begin{cases} x_1+x_3\geqslant 2,\\ x_2+2x_3\geqslant 5,\\ x_1,x_2,x_3\geqslant 0;\end{cases}$$

(2)
$$\min z=3x_1+2x_2+x_3+4x_4,$$
$$\begin{cases} 2x_1+4x_2+5x_3+x_4\geqslant 0,\\ 3x_1-x_2+7x_3-2x_4\geqslant 2,\\ 5x_1+2x_2+x_3+6x_4\geqslant 15,\\ x_1,x_2,x_3,x_4\geqslant 0.\end{cases}$$

2. 用对偶单纯形法求解下列线性规划问题：

(1)
$$\max 2x_1+x_2,$$
$$\begin{cases} 5x_2\leqslant 15,\\ 6x_1+2x_2\leqslant 24,\\ x_1+x_2\leqslant 5,\\ x_1,x_2\geqslant 0;\end{cases}$$

(2)
$$\max z=2x_1+4x_2+x_3+x_4,$$
$$\begin{cases} x_1+3x_2+x_4\leqslant 8,\\ 2x_1+x_2\leqslant 6,\\ x_2+x_3+x_4\leqslant 6,\\ x_1+x_2+x_3\leqslant 9,\\ x_j\geqslant 0(j=1,\cdots,4).\end{cases}$$

3. 现要做100套钢架，每套用长为2.9 m，2.1 m和1.5 m的元钢各一根. 已知原料长7.4 m，问应如何下料，使用的原材料最省？

4. 某饲养场饲养动物出售，设每头动物每天至少需700 g蛋白质、30 g矿物质、100 mg维生素. 现有物种饲料可供选用，各种饲料每千克营养成分含量及单价如下表所示：

饲料	蛋白质(g)	矿物质(g)	维生素(mg)	价格(元/千克)
1	3	1	0.5	0.2
2	2	0.5	1.0	0.7
3	1	0.2	0.2	0.4
4	6	2	2	0.3
5	18	0.5	0.8	0.8

要求确定既满足动物生长的营养需要，又使费用最省的选用饲料的方案？

5. 考虑线性规划问题

$$\max z = \alpha x_1 + 2x_2 + x_3 - 4x_4,$$

$$\text{s.t.}\begin{cases} x_1 + x_2 - x_4 = 4 + 2\beta, & (1)\\ 2x_1 - x_2 + 3x_3 - 2x_4 = 5 + 7\beta, & (2)\\ x_1, x_2, x_3, x_4 \geqslant 0. \end{cases}$$

模型中 α,β 为参数,要求:

(1) 组成两个新的约束(1)′=(1)+(2),(2)′=(2)-2(1),根据(1)′,(2)′,以 x_1,x_2 为基变量,列出初始单纯形表;

(2) 在表中,假设 $\beta=0$,则 α 为何值时,x_1,x_2 为问题的最优基;

(3) 在表中,假设 $\alpha=3$,则 β 为何值时,x_1,x_2 为问题的最优基;

6. 给出线性规划何题

$$\max z = x_1 + 2x_2 + x_3,$$

$$\begin{cases} x_1 + x_2 - x_3 \leqslant 2,\\ x_1 - x_2 + x_3 = 1,\\ 2x_1 + x_2 + x_3 \geqslant 2,\\ x_1 \geqslant 0, x_2 \leqslant 0, x_3 \text{ 无约束}. \end{cases}$$

(1) 写出其对偶问题;(2) 利用对偶问题性质证明原问题目标函数值 $z\leqslant 1$.

7. 写出下述线性规划的对偶规划

$$\max z = 6x_1 + 4x_2 + x_3 + 7x_4 + 5x_5,$$

$$(1)\begin{cases} 3x_1 + 7x_2 + 8x_3 + 5x_4 + x_5 = 3,\\ 2x_1 + x_2 + 3x_3 + 2x_4 + x_5 = 6,\\ x_1, x_2, x_3, x_4 \geqslant 0, x_5 \text{ 为自由变量}; \end{cases}$$

$$\min z = 2x_1 + 2x_2 + 4x_3,$$

$$(2)\begin{cases} 2x_1 + 3x_2 + 5x_3 \geqslant 2,\\ 2x_1 + x_2 + 7x_3 \leqslant 5,\\ x_1 + 4x_2 + 6x_3 \leqslant 5,\\ x_1, x_2 \geqslant 0, x_3 \text{ 为自由变量}. \end{cases}$$

8. 某工厂生产 A,B,C 三种产品,其所需劳动力、材料等有关数据见下表. 要求:(1)确定获利最大的产品生产计划;(2)产品 A 的利润在什么内变动时,上述最优计划不变;(3)如果设计一种新产品 D,单件劳动力消耗 8 单位,材料消耗为 2 单位,每件可获利 3 元,问该种产品是否值得生产?(4)如果劳动力数量不增加,材料不足时可从市场购买,每单位 0.4 元. 何该厂要不要购进原材料扩大生产,以购多少为宜?

消耗定额 资源 \ 产品	A	B	C	可用量(单位)
劳动力	6	3	5	45
材料	3	4	5	30
产品利润(元/件)	3	1	4	

9. 某服务部门各时段(每2个小时为一时段)需要的服务人数见下表.按规定,服务员连续工作8小时(即四个时段)为一班.现要求安排服务员的工作时间,使服务部门服务员总数最少?

时段	1	2	3	4	5	6	7	8
服务员人数	10	8	9	11	13	8	5	3

10. 用分枝定界法求解

(1) $$\max z = x_1 + x_2,\quad \begin{cases} x_1 + \dfrac{9}{14}x_2 \leqslant \dfrac{51}{14}, \\ -2x_1 + x_2 \leqslant \dfrac{1}{3}, \\ x_1, x_2 \geqslant 0, \\ x_1, x_2 \text{ 取整数}; \end{cases}$$

(2) $$\max z = 5x_1 - x_2 + 2x_3,\quad \begin{cases} 3x_1 + 10x_2 \leqslant 50, \\ 7x_1 - 2x_3 \leqslant 28, \\ x_1, x_2, x_3 \geqslant 0, \\ x_1, x_2, x_3 \text{ 取整数}. \end{cases}$$

11. 用割平面法求解下面整数规划

(1) $$\max z = 3x_1 - x_2,\quad \begin{cases} 3x_1 - 2x_2 \leqslant 3, \\ 5x_1 + 4x_2 \geqslant 10, \\ 2x_1 + x_2 \leqslant 5, \\ x_1, x_2 \geqslant 0, \\ x_1, x_2 \text{ 取整数}; \end{cases}$$

(2) $$\max z = 5x_1 + x_2,\quad \begin{cases} 3x_1 + x_2 \geqslant 9, \\ x_1 + x_2 \geqslant 5, \\ x_1 + 8x_2 \geqslant 8, \\ x_1, x_2 \geqslant 0, \\ x_1, x_2 \text{ 取整数}. \end{cases}$$

12. 解下列 0－1 型整数规划

(1) $$\min z = 5x_1 + 7x_2 + 10x_3 + 3x_4 + x_5,\quad \begin{cases} x_1 - 3x_2 + 5x_3 + x_4 - 4x_5 \geqslant 2, \\ -2x_1 + 6x_2 - 3x_3 - 2x_4 + 2x_5 \geqslant 0, \\ -2x_2 + 2x_3 - x_4 - x_5 \geqslant 1, \\ x_1, x_2, x_3, x_4, x_5 = 0 \text{ 或 } 1; \end{cases}$$

(2) $$\max z = 2x_1 + x_2 - x_3,\quad \begin{cases} x_1 + 3x_2 + x_3 \leqslant 2, \\ 4x_2 + x_3 \leqslant 5, \\ x_1 + 2x_2 - x_3 \leqslant 2, \\ x_1 + 4x_2 - x_3 \leqslant 4, \\ x_1, x_2, x_3 = 0 \text{ 或 } 1. \end{cases}$$

13. 解下列系数矩阵的最小化指派问题

$$(1)\begin{bmatrix}10 & 11 & 4 & 2 & 8\\ 7 & 11 & 10 & 14 & 12\\ 5 & 6 & 9 & 12 & 14\\ 13 & 15 & 11 & 10 & 7\end{bmatrix};\qquad (2)\begin{bmatrix}3 & 6 & 2 & 6\\ 7 & 1 & 4 & 4\\ 3 & 8 & 5 & 8\\ 6 & 4 & 3 & 7\\ 5 & 2 & 4 & 3\\ 5 & 7 & 6 & 2\end{bmatrix}.$$

14. 需要分派 5 人去做 5 项工作，每人做各项工作的能力评分见下表. 应如何分派，才能使总的得分最大？

业务 人员	B_1	B_2	B_3	B_4	B_5
A_1	1.3	0.8	0	0	1.0
A_2	0	1.2	1.3	1.3	0
A_3	1.0	0	0	1.2	0
A_4	0	1.05	0	0.2	1.4
A_5	1.0	0.9	0.6	0	1.1

第 12 章习题

1. 已知最优化问题的约束条件为

$$x_1+x_2+2x_3\leqslant 4,$$
$$x_1^2+x_2^2+x_3^2\leqslant 12,$$
$$x_1,x_2,x_3\geqslant 0.$$

分别写出下列各点的所有起作用约束：

(1) $x=(0,0,0)^T$, (2) $x=(2,2,0)^T$, (3) $x=(2\sqrt{3},0,0)^T$, (4) $x=\left(1,\frac{1}{2},\frac{1}{2}\right)^T$.

2. 求如下问题的 K－T 点.

(1)
$$\begin{aligned}\min\quad & (x_1-3)^2+(x_2-4)^2],\\ \text{s.t.}\quad & 16-(x_1-4)^2-x_2^2\geqslant 0,\\ & (x_1-3)^2+(x_2-2)^2=13,\end{aligned}$$

(2)
$$\begin{aligned}\min\quad & [(x_1-3)^2+(x_2-2)^2],\\ \text{s.t.}\quad & 5-x_1^2-x_2^2\geqslant 0,\\ & x_1\geqslant 0,\\ & x_2\geqslant 0,\\ & x_1+2x_2-4=0.\end{aligned}$$

3. 试说明约束问题

$$\min \quad [(x_1-3)^2+(x_2-4)^2],$$
$$\text{s.t. } 8-x_1^2+x_2^2 \geqslant 0,$$
$$x_1^2+x_2^2 \geqslant 0$$

在点 $x=(2,2)^T$ 处满足几何最优性条件,而在点 $x=(2,-2)^T$ 处不满足几何最优性条件.

4. 写出约束问题

$$\min \quad [(x_1-2)^2+(x_2-1)^2],$$
$$\text{s.t. } 9-x_1^2+x_2^2 \geqslant 0$$

的 K-T 条件.

5. 设问题

$$\min \quad (-x_1),$$
$$\text{s.t. } (1-x_1)^3-x_2 \geqslant 0,$$
$$x_1 \geqslant 0,$$
$$x_2 \geqslant 0.$$

(1) 写出与此问题对应的 K-T 条件;

(2) 检验 $x=(0,1)^T$ 和 $x=(1,0)^T$ 是否是 K-T 点,是否为最优点?

6. 设问题

$$\min \quad (-2x_1-3x_2),$$
$$\text{s.t. } x_1+x_2 \leqslant 8,$$
$$-x_1+2x_2 \leqslant 4,$$
$$x_1,x_2 \geqslant 0.$$

(1) 写出此问题的 K-T 条件;

(2) 验证最优点 $x^*=(4,4)^T$ 是 K-T 点;

(3) 验证如下的基本可行解 $x^{(1)}=(0,0)^T, x^{(2)}=(0,2)^T, x^{(3)}=(8,0)^T$ 都不是 K-T 点.

7. 设问题

$$\min f(x), x \in \mathbf{R}^n,$$
$$\text{s.t. } g_i(x) \geqslant 0, i=1,2,\cdots,m,$$

试证明,$\bar{x}$ 是如上问题的 K-T 点的充要条件为

$$\nabla f(\bar{x})^T d \geqslant 0,$$
$$\nabla g_i(\bar{x})^T d \geqslant 0, i \in I(x),$$

其中,d 为 $\bar{x}$ 处的一个可行方向.

8. 若 $f(x)$ 是可微凸函数,则 $\bar{x}$ 是规划问题

$$\min f(x),$$
$$x \geqslant 0$$

的最优解的充分必要条件为

$$\begin{cases} f_x(\bar{x}) \geqslant 0, \\ \bar{x} \geqslant 0, \\ f_x(\bar{x})\bar{x} = 0. \end{cases}$$

第 13 章习题

1. 试导出下列规划问题的鞍点问题

(1)(PP)
$$\begin{aligned} &\min f(\boldsymbol{x}), \\ &\text{s.t.}\quad \boldsymbol{g}(\boldsymbol{x}) \geqslant \boldsymbol{0}, \\ &\boldsymbol{x} \geqslant \boldsymbol{0}, \end{aligned}$$

其中 $\boldsymbol{x} \in \mathbf{R}^n, \boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \cdots, g_m(\boldsymbol{x}))^{\mathrm{T}}$.

(2)(PP)
$$\begin{aligned} &\min f(\boldsymbol{x}), \\ &\text{s.t.}\quad \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}, \end{aligned}$$

其中 $\boldsymbol{x} \in \mathbf{R}^n, \boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \cdots, g_m(\boldsymbol{x}))^{\mathrm{T}}$.

2. 试求下列规划问题的鞍点

(1)
$$\begin{aligned} &\min (x-1)^2 - 2x, \\ &\text{s.t.}\quad -x+1 \leqslant 0, \\ &x \in \mathbf{R}^1. \end{aligned}$$

(2)
$$\begin{aligned} &\min x_1 + x_2, \\ &\text{s.t.}\quad x_2 - x_1^3 \leqslant 0, \\ &x_2 \geqslant 0, \\ &x_1 \text{ 无限制}. \end{aligned}$$

3. 若有规划问题

$$(\text{PP})\quad \begin{aligned} &\min \quad \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\ &\text{s.t.}\quad \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b}, \\ &\boldsymbol{x} \geqslant \boldsymbol{0}, \end{aligned} \qquad (\text{DP})\quad \begin{aligned} &\max \quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{b}, \\ &\text{s.t.}\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A} \leqslant \boldsymbol{c}^{\mathrm{T}}, \\ &\boldsymbol{u} \geqslant \boldsymbol{0}. \end{aligned}$$

(1) 分别写出(PP)问题的 K－T 问题

(2) 若 $\bar{\boldsymbol{x}}$ 与 $\bar{\boldsymbol{u}}$ 分别为(PP)与(DP)问题的最优解，证明$(\bar{\boldsymbol{x}}, \bar{\boldsymbol{u}})$为(PP)的鞍点；反之亦真.

第 14、15、16 章习题

1. 试求约束问题

$$\begin{aligned} &\min f(\boldsymbol{x}) = (x_1-6)^2 + (x_2-2)^2, \\ &\text{s.t.}\quad -x_1 - 2x_2 \geqslant -4, \\ &-3x_1 - 2x_2 \geqslant -12, \\ &x_1, x_2 \geqslant 0 \end{aligned}$$

的下降可行方向.

2. 已知点 $\hat{\boldsymbol{x}}$ 满足问题

$$\min f(\boldsymbol{x}), \boldsymbol{x} \in \mathbf{R}^n,$$
$$\text{s.t.} \quad c_i(\boldsymbol{x}) \geq 0, i = 1, \cdots, m$$

的 K－T 条件，试证明对 $\boldsymbol{x}$ 处的任一可行方向 $\boldsymbol{d}$，都有成立下式

$$\nabla^{\mathrm{T}} f(\hat{\boldsymbol{x}}) \boldsymbol{d} \geq 0.$$

3. 用 Zoutendijk 可行方向法求解

(1)
$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = 4x_1^2 + x_2^2 - 32x_1 - 34x_2, \\ \text{s.t.} \quad & -x_1 \geq -2, \\ & -x_1 - 2x_2 \geq -6, \\ & x_1, x_2 \geq 0; \end{aligned}$$

(2)
$$\begin{aligned} \min \quad & \frac{4}{3}(x_1^2 - x_1 x_2 + x_2^2)^{3/4} - x_3, \\ \text{s.t.} \quad & -x_3 \geq -2, \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

(3)
$$\begin{aligned} \min \quad & (x_1^2 + 4x_2^2), \\ \text{s.t.} \quad & x_1 + x_2 \geq 1 (\boldsymbol{x}^{(1)} = (0,2)^{\mathrm{T}}), \\ & 15x_1 + 10x_2 \geq 12, \\ & x_1, x_2 \geq 0. \end{aligned}$$

4. 用 ε－起作用可行方向法求解

$$\begin{aligned} \min \quad & (x_1 - 2)^2 + (x_2 - 0.5)^2, \\ \text{s.t.} \quad & 1.5 - x_1^2 - x_2 \geq 0 (\boldsymbol{x}^{(1)} = (0,1)^{\mathrm{T}}), \\ & -x_1 + 2x_2 \geq 0, \\ & x_1 \geq 0. \end{aligned}$$

5. 用投影梯度法求问题

$$\begin{aligned} \min \quad & \{(x_1 - 3)^2 + (x_2 - 3)^2 + (x_3 - 1)^2\}, \\ \text{s.t.} \quad & 2x_1 + 3x_2 + 6x_3 \leq 6, \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

在点 $\boldsymbol{x}^{(1)} = (0,0,0)^{\mathrm{T}}$ 处的一个下降可行方向.

6. 用投影梯度法求解

(1)
$$\begin{aligned} \min \quad & x_1^2 + 2x_2^2 + (x_3 - 1)^2 + x_4, \\ \text{s.t.} \quad & x_3 \geq 0, \qquad (\boldsymbol{x}^{(1)} = (2,1,0,0)^{\mathrm{T}}) \end{aligned}$$

$$x_4 = 0.$$

(2)
$$\min\ 4x_1^2 + (x_2 - 2)^2,$$
$$\text{s.t.}\ x_1 \geqslant -2,$$
$$x_1 \leqslant 2, \qquad (x^{(1)} = (-1,1)^T)$$
$$x_2 \geqslant -1,$$
$$x_2 \leqslant 1.$$

(3)
$$\min\ \{2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2\},$$
$$\text{s.t.}\ -x_1 - x_2 \geqslant -2, \qquad (x^{(1)} = (0,0)^T)$$
$$-x_1 - 5x_2 \geqslant -5,$$
$$x_1 \geqslant 0, x_2 \geqslant 0.$$

7. 用外点法求解约束问题

(1)
$$\min\ f(x) = x_1^2 + x_2^2,$$
$$\text{s.t.}\ x_1 + x_2 - 2 = 0,$$

(2)
$$\min\ f(x) = \frac{1}{3}(x_1 + 1)^3 + x_2,$$
$$\text{s.t.}\ x_1 - 1 \geqslant 0,$$
$$x_2 \geqslant 0.$$

(3)
$$\min\ f(x_1, x_2) = x_1^2 + x_2^2,$$
$$\text{s.t.}\ 2x_1 + x_2 - 2 \leqslant 0,$$
$$-x_2 + 1 \leqslant 0.$$

8. 用障碍函数法求解

(1)
$$\min\ f(x) = x_1 + x_2,$$
$$\text{s.t.}\ x_1^2 + x_2^2 \geqslant 0,$$
$$x_1 \geqslant 0,$$

(2)
$$\min\ f(x) = \frac{1}{3}(x_1 + 1)^3 + x_2,$$
$$\text{s.t.}\ x_1 - 1 \geqslant 0,$$
$$x_2 \geqslant 0.$$

并与 7 题中的(2)作比较.

9. 用广义乘子法求解

(1)
$$\min \quad (x_1^2 + 2x_2^2)(\text{取 } v_1 = 0),$$
$$\text{s.t.} \quad x_1 + x_2 - 1 = 0.$$

(2)
$$\min \quad f(x) = x_1^2 + x_1x_2 + x_2^2,$$
$$\text{s.t.} \quad x_1 + 2x_2 = 4.$$

10. 写出如下二次规划问题的 K－T 条件

$$\min \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\boldsymbol{x} + (1,1,1)\boldsymbol{x},$$

$$\text{s.t.} \quad x_1 + 2x_2 + x_3 = 4.$$

11. 求解二次规划问题

$$\min \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\begin{bmatrix} 3 & -1 & 2 \\ -1 & 2 & 0 \\ 2 & 0 & 4 \end{bmatrix}\boldsymbol{x} + (1, -3, -2)\boldsymbol{x},$$

$$3x_1 - 2x_2 + 5x_3 \leqslant 4,$$
$$-2x_1 + 3x_2 + 2x_3 \leqslant 3,$$
$$x_1, x_2, x_3 \geqslant 0.$$

第 17 章习题

1. 机器负荷分配问题

某公司有资金 10 万元，若投资于项目 $i(i=1,2,3)$ 的投资为 x_i 时，其收益分别为 $g_1(x_1)=4x_1, g_2(x_2)=9x_2, g_3(x_3)=2x_3^2$，问应如何分配投资额才能使总收益最大？

2. 有一辆最大货运量为 10 吨的卡车，用以装载 3 种货物，每种货物的单位重量级相应单位价值如下表所示. 应如何装载可使总价值最大？

货物编号(i)	1	2	3
单位重量(t)	3	4	5
单位价值 c_i	4	5	6

3. 用逆推解法和顺推解法求解下面问题

$$\max z = x_1 \times x_2^2 \times x_3,$$

$$\begin{cases} x_1 + x_2 + x_3 = c(c > 0), \\ x_i \geqslant 0, i = 1,2,3. \end{cases}$$

4. 生产与存贮问题

某工厂生产并销售某种产品，已知今后四个月市场需求预测如下表，由每月生产 j 单位产品费用为：

$$C(j) = \begin{cases} 0 & (j = 0), \\ 3 + j & (j = 1, \cdots, 6)\text{（千元）}. \end{cases}$$

每月库存单位产品的费用为，该厂最大库存容量为 3 单位，每月最大生产能力为 6 单位，计划开始和计划期末库存量都是零. 试制定四个月的生产计划，在满足用户需求条件下总费用最小.

i（月）	1	2	3	4
g_i（需求）	2	3	2	4

5. 设某人有 400 万元现金，计划在四年内若投资用去万元就能获得万元的效用. 每年没有用掉的金额，连同利息（年利息 10%）可再用余下一年的投资. 而每年以打算用于投资的金额不计利息. 是指定金额的使用计划，而使四年内获得的总效用最大？

（1）用动态规划方法求解；（2）用拉格朗日乘数法求解；（3）比较两种解法，并说明动态规划方法有哪些优点？

6. 试用动态规划的方法求解

$$\max f(X) = 4x_1 + x_2 + x_3,$$

$$\begin{cases} x_1 + x_2 \leqslant 1, \\ 2x_1 + x_3 \leqslant 2, \\ x_1, x_2, x_3 \text{ 为非负整数}. \end{cases}$$

7. 设有一个工厂为市场生产某种产品，每月月初对产品的销售情况进行一次检查，其结果有二：销售好（记为状态 1）；也可能销售差（记为状态 2）. 若处于状态 1，由于各种随机因素的干扰，下月初仍处于销路好的概率为 0.5，转为销路差的概率也为 0.5；若处于状态 2，则下月初转为销路好的概率为 0.4，仍处于销路差的概率为0.6. 这状态转移矩阵为

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{bmatrix}.$$

若某月初销路好，下月初仍销路好可获利 9 千元，下月初转为销路差可获利 3 千元；若某月初销路差，下月初转为销路好可获利 3 千元，下月初转为销路差要亏本 7 千元. 则报酬矩阵为

$$\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} 9 & 3 \\ 3 & -7 \end{bmatrix}.$$

问题是在若干月内采取什么决策才能使其总期望报酬最大.

8. 某公司打算在3个不同地区设置4个销售点，根据市场预测部门估计，在不同地区设备不同数量的销售站，每月可得利润如下表所示，试问应如何在各地区设置销售站，可使每月总利润最大？

地区	销售站				
	0	1	2	3	4
1	0	16	25	30	32
2	0	12	17	20	22
3	0	10	14	16	17

9. 为保证某设备正常运转，需对串联工作的三种不同零件 A_1, A_2, A_3，分别确定备件数量. 若增加备用零件的数量，可提高设备正常运转的可靠性，但费用要增加，而总投资额为8千元. 已知备用零件数与它的可靠性和费用关系如下表所示，求 A_1，A_2, A_3，A 得备用零件数量各为多少时，可使设备运转的可靠性最高.

备件数	可靠性			备用零件费用(千元)		
	A_1	A_2	A_3	A_1	A_2	A_3
1	0.3	0.2	0.1	1	3	2
2	0.4	0.5	0.2	2	5	3
3	0.5	0.9	0.7	3	6	4

参 考 文 献

[1] Wolf M. A.. Numerical Methods for Unconstrained optimization An Metroduction. Van Nostrand Reinbold Company, 1978

[2] Mokhtar. S. Bazaraa, C. M. Shetty Nonlinear Programming Theory and Algorithms. John Wiley and Sons. 1979

[3] Hestenes M. R.. E. Stiefel. Methods of conjugate Gradients for Solving linear systems. J. Res. Nat. Bur. Stand, 1972:49

[4] L. C. W. Dixon, Conjugate Gradient Algorithms Quadratic Termination without linear Searches, J. Inst. Maths. Applies, 1975:15

[5] Zangwill, W. I. Management Science 13, 1967a: 344—358. Nonliner Programming Via Penalty functions

[6] 席少霖.非线性最优化方法.北京:高等教育出版社,1992

[7] 赵凤治,尉继英.约束最优化计算方法.北京:科学出版社,1991

[8] 王德人.非线性方程组解法与最优化方法.北京:人民教育出版社,1979

[9] 魏权龄等.数学规划与优化设计.北京:国防工业出版社,1984

[10] 张光澄等.最优化方法.成都:成都科技大学出版社,1990

[11] Auriel M.李元嘉等译.非线性规划——分析与方法上、下册.上海:上海科技出版社,1979

[12] Himmelblan D. M.. Applied Nonlinear Programming. 1972(中译本,实用非线性规划.北京:科学出版社,1983)

[13] 邢文训,谢金星.现代优化计算方法.北京:清华大学出版社,2003

[14] Bazaraa N. S., Shetty C. M. Nonliner Programming: Theory and Algorithms. John Wiley and Sons, 1997: 142—143

[15] 薛嘉庆.最优化原理与方法.北京:冶金工业出版社,1983

[16] 谢政,李建平,汤泽滢.非线性最优化.长沙:国防科技大学出版社,2003

[17] 杨浩.模型与算法.北京:北方交通大学出版社,2002

[18] 《运筹学》教材编写组.运筹学.北京:清华大学出版社,1990

[19] 张莹.运筹学基础.北京:清华大学出版社,1995

[20] 胡运权,郭耀煌.运筹学教程.北京:清华大学出版社,1998

[21] 程理民,吴江,张玉林.运筹学模型与方法教程.北京:清华大学出版社,2000
[22] 杨建刚.人工神经网络实用教程.杭州:浙江大学出版社,2001
[23] 阎平凡,张长水.人工神经网络与模拟进化计算.北京:清华大学出版社,2000
[24] 邓乃扬.无约束最优化计算方法.北京:科学出版社,1982
[25] 何旭初,孙麟平.约束最优化方法.南京:南京大学出版社,1986
[26] 薛毅.最优化原理与方法.北京:北京工业大学出版社,2001